2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
ASEE
Paper ID #39050

An Investigation of the Effect of Number of Hot Spots on Taxi Time at U.S. Hub Airports

Mr. Shantanu Gupta, Purdue University at West Lafayette (PPI)

Shantanu Gupta is a PhD candidate in the School of Aviation and Transportation Technology at Purdue University with Dr. Mary E. Johnson. He earned his B.E in Mechanical Engineering from Visvesvaraya Technological University, India, and M.S in Aviation and Aerospace Management from Purdue University, West Lafayette. Mr. Gupta is currently working with Dr. Johnson on the PEGASAS Project 33 – Augmented Weather Information Project (AWIP) as research assistant.

Prof. Mary E. Johnson Ph.D., Purdue University at West Lafayette (PPI)

Mary E. Johnson is a Professor and Associate Head for Graduate Studies in the School of Aviation and Transportation Technology (SATT) at Purdue University in West Lafayette, Indiana. She earned her BS, MS and PhD in Industrial Engineering from The University of Texas at Arlington. After 5 years in aerospace manufacturing as an IE, Dr. Johnson joined the Automation & Robotics Research Institute in Fort Worth and was a program manager for applied research programs. Fourteen years later, she was an Industrial Engineering assistant professor at Texas A&M - Commerce before joining the Aviation Technology department at Purdue University in West Lafayette, Indiana in 2007 as an Associate Professor. She is a Co-PI on the FAA Center of Excellence for general aviation research known as PEGASAS and leads the Graduate Programs in SATT. Her research interests are aviation sustainability, data driven process improvement, and aviation education.

Mr. Jiansen Wang, Purdue University

Jiansen is a PhD student in the School of Aviation and Transportation Technology at Purdue University. He began his PhD study in Aviation and Aerospace Management at Purdue University in 2021, under the supervision of Dr. Mary E Johnson. His research focuses on optimizing airport sustainability. Jiansen completed his M.S degree in Aviation and Aerospace Management at Purdue University in 2020. During his masterÄôs study, Jiansen earned second prize in Airport Cooperative Research Program Competition in 2020. Prior to graduate school, Jiansen completed his B.S. degree in Engineering from Civil Aviation University of China.

©American Society for Engineering Education, 2023

An Investigation of the Effect of Number of Hot Spots on Taxi-Time at U.S Hub Airports

INTRODUCTION

Hotspots on an airport movement area may require heightened attention by pilots and controllers, which may affect taxi times at airports. Taxi time could affect airport congestion, engine emissions related to air pollutants, and aircraft fuel consumption. Airport congestion affects airport capacity and aircraft fuel burn. Aircraft operations, including taxi operations, contribute to fuel consumption and engine exhaust emissions at airports [1]. When taxiing, the fuel efficiency of stop-and-go situations is 35% higher than that in unimpeded situations [2].

Hotspots are areas that have a history or potential risk of collisions or runway incursions [3]. In general, hotspots are complex or confusing taxiway/taxiway or taxiway/runway intersections at an airport, which are identified and depicted on the respective airport diagrams by the Federal Aviation Administration (FAA)[4]. In this paper, the researchers aim to better understand taxi time at airports, and the potential effect of the number of airport hotspots on the taxi time at these airports. This research aims to find whether taxi time at airports differ by airport hub classifications and by the number of hot spots on airports.

For this study, a sample of 33 airports was selected from the 77 airports listed in the Aviation System Performance Metrics (ASPM) [5] data published by the FAA. The researchers sampled the 11 busiest airports (by number of operations conducted) from each of the three hub categories – Large (L), Medium (M), and Small (S) – as identified by the National Plan of Integrated Airport Systems (NPIAS) [6]. The 20 busiest days (by number of operations conducted) from May 01, 2022, to September 30, 2022 were selected for each airport. From the ASPM dataset, average quarter-hour taxi-in and taxi-out times between 06:00AM to 10:00 PM were collected for each of the airports and their 20 busiest days, respectively. The researchers used FAA published airport diagrams (26 January 2023 to 23 February 2023) to count the number of hot spots for each of the 33 airports. Statistical and graphical tests were used to answer the research questions.

This study may help in better understanding and modelling the taxi times that can be used to reduce congestion, fuel burn, and emissions at airports. This may potentially increase airport capacity to meet the increasing traffic demand. The results of this study may be used to teach airport planning, operations, and real-world statistical analyses in engineering and technology courses. This research paper may have practical applications in statistical analyses and discrete-event stochastic process simulation. This paper uses parametric and non-parametric statistical tests to answer research questions, and a narrative approach for data analysis is followed so that the instructors and students may follow along with the thought process. Instructors may be able to use this paper to highlight research methodology and findings when working with real world data, assumptions of common statistical methods fail, and there is an abundance of datapoints.

BACKGROUND

In airports with air traffic controllers (ATCs), the ATCs give taxi clearance for pilots to follow while the aircraft is in the airport movement area. Given the capacity and workload of the airport, the aircraft pilots may be given a route that is not fuel and/or time efficient. During busy hours, the aircraft may experience delay due to frequent braking to avoid other aircraft or ground vehicles. Taxi time refers to the time differences between the actual gate time and the wheel time. Taxi time is related to the aircraft fuel consumption, airport congestion, and engine emissions of air pollutants at the airport. When taxiing, stop-and-go situations account for about 18% of fuel consumed, which is approximately 35% higher than operating aircraft in unimpeded situations [2]. Congestion on airport surfaces is a significant constraint to the available capacity of the air transportation system [7]. By proposing a new sequential graph-based algorithm to optimize routing in Zurich airport, an estimated average of 136.9 seconds may be reduced on taxi time per aircraft, therefore saving an estimated total of $9.6 million on fuel cost per year [8]. This research focuses on analyzing the relationship between number of hot spots and quarter-hour average taxi time in small, medium, and large hub airports in the U.S. Investigating and analyzing the relationship between the number of hot spots and taxi time may reduce the taxi time; therefore, mitigate congestion, reduce fuel burn and engine exhaust emissions at airports.

NPIAS airports & Hub Classification: The FAA classifies public-use airports as commercial service, reliever, and general aviation airports [9]. The commercial service airports are further classified as large hub, medium hub, small hub, and non-hub airports [9]. The FAA defines large hub airports that receive more than 1 % of the annual U.S. commercial enplanements, medium hubs as airports that receive 0.25% to 1% annual enplanements, small hubs as airports that receive 0.05% to 0.25% annual enplanements, and non-hubs as airports that receive less than 0.05% but more than 10,000 annual enplanements [9]. The National Plan of Integrated Airport System (NPIAS) identifies approximately 3300 public-use airports and assesses their eligibility for Federal funding every two years [6]. In the NPIAS report, the "Appendix A: List of NPIAS Airports" contains information of the airports that are documented in the corresponding NPIAS reports including the airport hub classification [10].

ASPM and Taxi time: The FAA Aviation System Performance Metrics (ASPM) dataset tracks, collects, and reports on the operation and performance data of airport and airlines [11]. The dataset publishes performance data from both arrival and departure operations in the 77 ASPM airports, and ASPM airlines [11]. The taxi-in time refers to the average difference between actual gate time and actual wheels on time, in minutes [12]. The taxi-out time refers to the average difference between Actual Wheels Off time and Actual Gate Out time, in minutes [12].

Hotspots: A hot spot is an airport movement area that may require heighted attention by pilots and air traffic controllers, which may affect taxi times at airports [4]. The FAA publishes data of public airports on the Digital – Chart Supplement (d-CS) every 56 days, which includes the number of hot spots at public-use airports [4].

RESEARCH QUESTIONS

In this paper, the researchers aim to better understand taxi time at small, medium, and large hub airports, and the potential effect of the number of airport hotspots on the taxi time at these airports. Specifically, this research aims to answer these research questions:
RQ1: Does taxi time differ by airport hub classifications?
RQ2: Does taxi time differ by the number of hot spots on airports?
RQ3: Does taxi time differ by the number of hot spots on different airport hub classifications?

METHODOLOGY

This section describes the data sources, collection, consolidation, and analyses conducted to answer the research questions.

Data Sources and Collection

For this study, FAA published airport-related data was collected from ASPM (Airport Analysis [12] and Taxi Times [13]), NPIAS 2023-2027 (Hub Classification [10]), and FAA airport diagrams [14]. Data collection, selection, and sampling was conducted as follows:

1. **ASPM 77 [12]** dataset was used for the list of airports, dates, and the number of daily Departures and Arrivals for Metric Computation. The researchers decided to capture a summer travel time frame and therefore, selected a time-frame between 05/01/2022 and 09/31/2022. Within the ASPM, the data was grouped by Airports and by Dates to run the query. Figure 1 shows a snippet of the query run on ASPM dataset and a section of the resulting MS Excel worksheet.

Figure 1. Snapshot of the ASPM [12] query and part of the downloaded data.

2. **NPIAS 2023-2027 [10]** dataset was used to find the hub classification of airports. The dataset obtained from ASPM contained information about 77 airports. The NPIAS dataset was crossmatched with the ASPM dataset to find the hub classification (Small/Medium/Large) of the 77 airports.

3. **To find the busiest airports,** total sum of number of Departures and Arrivals for all days from 05/01/2022 to 09/30/2022 was calculated for each airport. Then, airports in each of the hub classifications (S/M/L) were sorted (from largest to smallest) based on the sum of departures and arrivals. The researchers selected 11 busiest airports in each of the three hub classifications to form a sample of 33 airports.

4. **To find the busiest days for each airport,** sum of departures and arrivals for each day was calculated from 05/01/2022 to 09/30/2022 for each airport. Then, dates for each airport were sorted (from largest to smallest) based on the sum of daily departures and arrivals. The researchers selected 20 busiest days for each of the 33 airports. Note: the busiest days may differ for each of the airports

5. **To find the taxi-time data**, quarter hour taxi-time data was collected from ASPM dataset [13] for each of the 20 busiest days (step 4) for each of the 33 airports (step 3). Local time from 6:00 AM to 10:00 PM selected.

6. **To find the number of hotspots** for each airport, FAA Airport Diagrams [14] (26 January 2023 to 23 February 2023) of each of the 33 airports were investigated. Number of hotspots (as identified and reported by the airports on their airport diagrams) were counted and noted for each airport.

Data Consolidation

The data collected for the 33 airports from ASPM, NPIAS, and airport diagrams were consolidated into one MS Excel worksheet. For each of the 33 airports, quarter-hour taxi time (in and out) between 6:00 AM to 10:00 PM (local time) for 20 busiest days, number of departures and arrivals for metric computation in each quarter, NPIAS hub classification, and number of hotspots were tabulated. The researchers collected 39,268 observations across 10 fields. Figure 2 shows a snippet from the table.

Figure 2. Snapshot of the consolidated data table using three sources of data.

ASPM Dataset [12]								NPIAS 2023-2027 [10]	FAA Airport Diagrams [14]
Facility	Date	Quarter	Hour	Departures For Metric Computation	Average taxi out time	Arrivals For Metric Computation	Average taxi in time	NPIAS Hub Classification	Number of Hot Spots
STL	09/20/2022	4	15	2	10	4	5	M	1
STL	09/20/2022	4	16	2	11	7	4.86	M	1
STL	09/20/2022	4	17	2	12	4	6.5	M	1
STL	09/20/2022	4	18	4	11.5	5	3.8	M	1
STL	09/20/2022	4	19	5	10.8	1	5	M	1
STL	09/20/2022	4	20	1	12	12	6.92	M	1
STL	09/20/2022	4	21	2	10	2	5.5	M	1
ATL	5/5/2022	1	6	10	17.3	2	12.5	L	2
ATL	5/5/2022	1	7	8	14	5	5.8	L	2
ATL	5/5/2022	1	8	21	14.86	15	8.13	L	2
ATL	5/5/2022	1	9	20	16.85	29	9.59	L	2
ATL	5/5/2022	1	10	42	21.24	18	7.67	L	2
ATL	5/5/2022	1	11	33	13.64	11	6.09	L	2

Data Analysis

The researchers collected 39,268 observations across 10 fields to answer the research questions. The researchers applied specific parametric and non-parametric tests to answer these questions. It is important to note that the data analysis follows a narrative approach so that instructors and students may follow along the thought processes.

The researchers approached each of the research questions in two parts to test taxi-out time and taxi-in time separately. One-way analysis of variance (ANOVA) was used to compare the taxi-time means by airport hub classification (S/M/L), number of hotspots (0, 1, 2, 3, 4, 5), and by number of hotspots on each of the hub classifications. The researchers explored both methods of ANOVA – assuming equal variances followed by Tukey post hoc test, and not assuming equal variances (Welch's test) followed by Games-Howell post hoc test. Since the data indicated the presence of numerous outliers, the researchers also applied the Kruskal-Wallis non-parametric test to check median taxi times across hub classifications and number of hotspots. These tests were repeated for taxi-out and taxi-in times.

Research Question 1

For the one-way ANOVAs (and the Welch's tests), the null and alternate hypotheses to test mean taxi-in and taxi-out times by NPIAS hub classifications were:

H_o: $\mu_S = \mu_M = \mu_L$ (for both taxi-in and taxi-out times)

H_a: mean taxi-out (or taxi-in) time is different for at least one hub classification

For the Kruskal-Wallis test, the null and alternate hypotheses to test median taxi-in and taxi-out times by NPIAS hub classifications were:

H_o: $\eta_S = \eta_M = \eta_L$ (for both taxi-in and taxi-out times)

H_a: median taxi-out (or taxi-in) time is different for at least one hub classification

Research Question 2

For the one-way ANOVAs (and the Welch's tests), the null and alternate hypotheses to test mean taxi-in and taxi-out times by number of hotspots were:

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times)

H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Whereas, for the Kruskal-Wallis test, the null and alternate hypotheses to test median taxi-in and taxi-out times by number of hotspots were:

H_o: $\eta_0 = \eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5$ (for both taxi-in and taxi-out times)

H_a: median taxi-out (or taxi-in) time is different for at least one hotspot count

Research Question 3

For the one-way ANOVAs (and the Welch's tests), the null and alternate hypotheses to test mean taxi-in and taxi-out times by number of hotspots on small, medium, and large hubs were:

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times at S/M/L hubs)

H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Whereas, for the Kruskal-Wallis test, the null and alternate hypotheses to test median taxi-in and taxi-out times by number of hotspots on small, medium, and large hubs were:

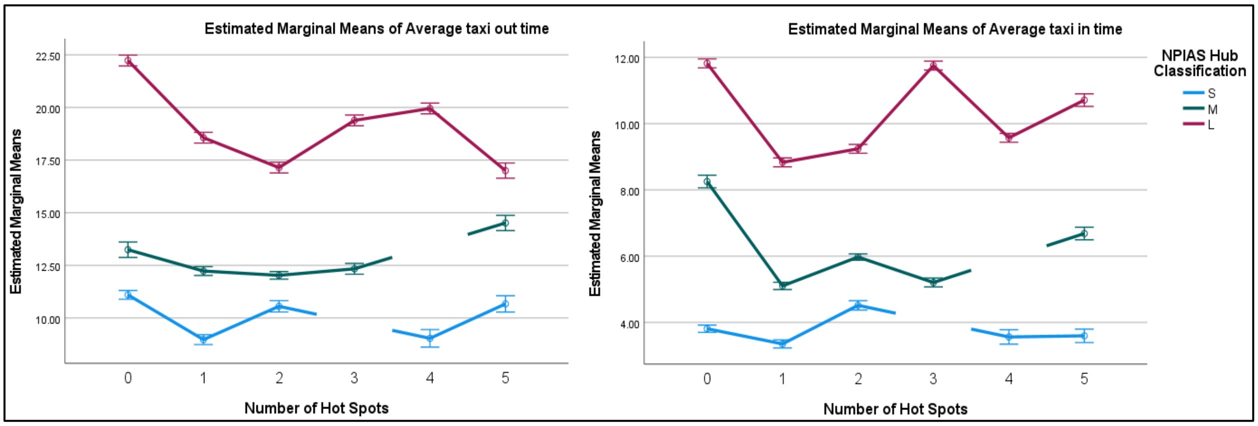
H_o: $\eta_0 = \eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5$ (for both taxi-in and taxi-out times at S/M/L hubs)

H_a: median taxi-out (or taxi-in) time is different for at least one hotspot count

RESULTS

This section presents the results of the statistical tests used to answer the research questions. Through the research question, the researchers aimed to study taxi times (taxi-out and taxi-in times) by airport hub classification, number of hotspots on airports, and by number of hotspots on specific hub classifications. Figure 3 demonstrates the mean taxi-out and taxi-in times for small, medium, and large hubs for different number of hotspots. Detailed results and snapshots of the statistical tests are shown in the appendix.

Figure 3. Mean taxi-in and taxi-out time by hub classification and number of hotspots.



RQ1. Does taxi time differ by airport hub classifications? Refer Table 1 for detailed statistics.

<u>One-way ANOVA:</u> Using the data collected, an alpha of 0.05, and one-way ANOVA (assuming equal variances) test to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across NPIAS hub classifications. Using the Tukey HSD post hoc and 95% confidence, the mean taxi-out (and taxi-in) time was found to be different for each of the hub classifications – small, medium, and large.

<u>Welch's Test:</u> Using the data collected, an alpha of 0.05, and one-way ANOVA (not assuming equal variances) test to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across NPIAS hub classifications. Using Games-Howell post hoc and 95% confidence, the mean taxi-out (and taxi-in) time was found to be different for each of the hub classifications – small, medium, and large.

<u>Kruskal-Wallis Test:</u> Using the data collected, an alpha of 0.05, and the Kruskal-Wallis non-parametric test to compare medians, the researchers *rejected the null hypotheses* (p-value<0.001) that the median taxi-out (and taxi-in) time was same across NPIAS hub classifications. Therefore, the researchers concluded that median taxi-out (and taxi-in) time was different for at least one hub classification. No post hoc tests were conducted. The mean ranks and Z-value indicate that the taxi-out (and taxi-in) times at large hub airports tend to be higher than those at medium hubs and small hubs airports. Mean rank of large hub airport was greater than the overall mean rank, whereas, mean ranks of medium and small hub airports were less than the overall mean rank.

Table 1. Results of RQ1

RQ1 — Taxi time vs Hubs

Taxi-out Time

Means

NPIAS Hub Classification	N	Mean	StDev	95% CI
L	14078	19.2336	6.3782	(19.1228, 19.3443)
M	13970	12.4763	5.4922	(12.3651, 12.5875)
S	11220	10.1944	8.2902	(10.0704, 10.3185)

Pooled StDev = 6.70474

Taxi-in Time

Means

NPIAS Hub Classification	N	Mean	StDev	95% CI
L	14078	10.2852	3.7082	(10.2261, 10.3443)
M	13970	5.8661	3.2951	(5.8068, 5.9254)
S	11220	3.7904	3.7423	(3.7242, 3.8566)

Pooled StDev = 3.57691

One-way ANOVA (Assuming equal variances) — Post hoc: Tukey HSD

H_o: $\mu_S = \mu_M = \mu_L$ (for both taxi-in and taxi-out times)
H_a: mean taxi-out (or taxi-in) time is different for at least one hub classification

Taxi-out Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
NPIAS Hub Classification	2	578132	289066	6430.33	0.000
Error	39265	1765099	45		
Total	39267	2343231			

Grouping Information Using the Tukey Method and 95% Confidence

NPIAS Hub Classification	N	Mean	Grouping
L	14078	19.2336	A
M	13970	12.4763	B
S	11220	10.1944	C

Means that do not share a letter are significantly different.

Taxi-in Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
NPIAS Hub Classification	2	284681	142340	11125.31	0.000
Error	39265	502368	13		
Total	39267	787049			

Grouping Information Using the Tukey Method and 95% Confidence

NPIAS Hub Classification	N	Mean	Grouping
L	14078	10.2852	A
M	13970	5.8661	B
S	11220	3.7904	C

Means that do not share a letter are significantly different.

Welch's Test (Not assuming equal variances) — Post hoc: Games-Howell

H_o: $\mu_S = \mu_M = \mu_L$ (for both taxi-in and taxi-out times)
H_a: mean taxi-out (or taxi-in) time is different for at least one hub classification

Taxi-out Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
NPIAS Hub Classification	2	24060.9	6338.31	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

NPIAS Hub Classification	N	Mean	Grouping
L	14078	19.2336	A
M	13970	12.4763	B
S	11220	10.1944	C

Means that do not share a letter are significantly different.

Taxi-in Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
NPIAS Hub Classification	2	25216.6	10406.44	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

NPIAS Hub Classification	N	Mean	Grouping
L	14078	10.2852	A
M	13970	5.8661	B
S	11220	3.7904	C

Means that do not share a letter are significantly different.

Kruskal-Wallis (non-parametric)

H_o: $\eta_S = \eta_M = \eta_L$ (for both taxi-in and taxi-out times)
H_a: median taxi-out (or taxi-in) time is different for at least one hub classification

Taxi-out Time

Descriptive Statistics

NPIAS Hub Classification	N	Median	Mean Rank	Z-Value
L	14078	17.75	28448.6	115.19
M	13970	12.40	15767.9	-50.23
S	11220	11.00	13389.5	-69.05
Overall	39268		19634.5	

Test

Null hypothesis H_0: All medians are equal
Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	2	13541.83	0.000
Adjusted for ties	2	13562.44	0.000

Taxi-in Time

Descriptive Statistics

NPIAS Hub Classification	N	Median	Mean Rank	Z-Value
L	14078	9.63	29676.3	131.23
M	13970	5.40	16745.2	-37.53
S	11220	4.00	10632.2	-99.53
Overall	39268		19634.5	

Test

Null hypothesis H_0: All medians are equal
Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	2	19031.06	0.000
Adjusted for ties	2	19074.12	0.000

RQ2. Does taxi time differ by the number of hot spots on airports? Refer to Table 2 for detailed statistics.

One-way ANOVA: Using the data collected, an alpha of 0.05, and one-way ANOVA (assuming equal variances) test to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across the number of hotspots.

> *Taxi-out time:* Using the Tukey HSD post hoc and 95% confidence, the mean taxi-out time was found to be significantly different for airports with 0, 3, 4, and 5 hotspots. There was no significant difference in mean taxi-out time for airports with 1 or 2 hotspots, but they differed collectively from airports with 0, 3, 4, or 5 hotspots.
>
> *Taxi-in time:* Using the Tukey HSD post hoc and 95% confidence, the mean taxi-in time was found to be significantly different for airports with 1, 2, 3, and 4 hotspots. There was no significant statistical difference in mean taxi-in time for airports with 0 or 5 hotspots, but they differed collectively from airports with 1, 2, 3, or 4 hotspots.

Welch's Test: Using the data collected, alpha of 0.05, and one-way ANOVA (not assuming equal variances) test to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across the number of hotspots.

> *Taxi-out time:* Using the Games-Howell post hoc and 95% confidence, the mean taxi-out time was found to be significantly different for airports with 0, 3, 4, and 5 hotspots. There was no significant difference in mean taxi-out time for airports with 1 or 2 hotspots, but they differed collectively from airports with 0, 3, 4, and 5 hotspots.
>
> *Taxi-in time:* Using the Games-Howell post hoc and 95% confidence, the mean taxi-in time was found to be significantly different for airports with 1, 2, 3, or 4 hotspots. There was no significant difference in mean taxi-in time for airports with 0 or 5 hotspots, but they differed collectively from airports with 1, 2, 3, or 4 hotspots.

Kruskal-Wallis Test: Using the data collected, an alpha of 0.05, and the Kruskal-Wallis non-parametric test to compare medians, the researchers *rejected the null hypotheses* (p-value<0.001) that the median taxi-out (and taxi-in) time was same across the number of hotspots. Therefore, the researchers concluded that median taxi-out (and taxi-in) time was different for at least one hotspot count. No post hoc tests were conducted. The mean ranks and Z-value indicate that the taxi-out (and taxi-in) times at airports with 0, 3, 4, and 5 hotspots tend to be higher than those at airports with 1 and 2 hotspots. Mean rank of airports with 0, 3, 4, and 5 hotspots was greater than the overall mean rank, whereas, mean ranks of airports with 1 and 2 hotspots were less than the overall mean rank.

Table 2. Results of RQ2

RQ2 — Taxi time vs Hotspots

Taxi-out Time

Means

Number of Hot Spots	N	Mean	StDev	95% CI
0	7632	15.189	9.872	(15.019, 15.360)
1	9425	12.9044	7.7205	(12.7511, 13.0578)
2	9907	13.0075	6.3964	(12.8580, 13.1571)
3	5115	15.8632	6.0948	(15.6551, 16.0714)
4	3523	16.968	7.778	(16.717, 17.219)
5	3666	14.214	6.419	(13.969, 14.460)

Pooled StDev = 7.59500

Taxi-in Time

Means

Number of Hot Spots	N	Mean	StDev	95% CI
0	7632	7.2324	5.5167	(7.1343, 7.3305)
1	9425	5.5541	3.8720	(5.4658, 5.6424)
2	9907	6.4804	3.6620	(6.3943, 6.5665)
3	5115	8.4795	4.4751	(8.3596, 8.5993)
4	3523	7.9301	4.8885	(7.7857, 8.0745)
5	3666	7.1513	3.9706	(7.0097, 7.2928)

Pooled StDev = 4.37276

One-way ANOVA (Assuming equal variances) — Post hoc: Tukey HSD

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times)

H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Number of Hot Spots	5	78439	15687.8	271.96	0.000
Error	39262	2264792	57.7		
Total	39267	2343231			

Grouping Information Using the Tukey Method and 95% Confidence

Number of Hot Spots	N	Mean	Grouping
4	3523	16.968	A
3	5115	15.8632	B
0	7632	15.189	C
5	3666	14.214	D
2	9907	13.0075	E
1	9425	12.9044	E

Means that do not share a letter are significantly different.

Taxi-in Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Number of Hot Spots	5	36319	7263.83	379.89	0.000
Error	39262	750730	19.12		
Total	39267	787049			

Grouping Information Using the Tukey Method and 95% Confidence

Number of Hot Spots	N	Mean	Grouping
3	5115	8.4795	A
4	3523	7.9301	B
0	7632	7.2324	C
5	3666	7.1513	C
2	9907	6.4804	D
1	9425	5.5541	E

Means that do not share a letter are significantly different.

Welch's Test (Not assuming equal variances) — Post hoc: Games-Howell

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times)

H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Number of Hot Spots	5	14476.2	302.50	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

Number of Hot Spots	N	Mean	Grouping
4	3523	16.968	A
3	5115	15.8632	B
0	7632	15.189	C
5	3666	14.214	D
2	9907	13.0075	E
1	9425	12.9044	E

Means that do not share a letter are significantly different.

Taxi-in Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Number of Hot Spots	5	14133.6	402.19	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

Number of Hot Spots	N	Mean	Grouping
3	5115	8.4795	A
4	3523	7.9301	B
0	7632	7.2324	C
5	3666	7.1513	C
2	9907	6.4804	D
1	9425	5.5541	E

Means that do not share a letter are significantly different.

Kruskal-Wallis (non-parametric)

H_o: $\eta_0 = \eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5$ (for both taxi-in and taxi-out times)

H_a: median taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Descriptive Statistics

Number of Hot Spots	N	Median	Mean Rank	Z-Value
0	7632	15.00	20920.1	11.04
1	9425	12.50	16505.2	-30.74
2	9907	13.50	17907.6	-17.54
3	5115	15.40	22501.6	19.40
4	3523	17.14	25035.3	29.64
5	3666	14.81	20479.9	4.74
Overall	39268		19634.5	

Test

Null hypothesis H_0: All medians are equal

Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	5	2193.61	0.000
Adjusted for ties	5	2196.94	0.000

Taxi-in Time

Descriptive Statistics

Number of Hot Spots	N	Median	Mean Rank	Z-Value
0	7632	6.50	20135.5	4.30
1	9425	5.00	15825.9	-37.42
2	9907	6.00	19106.2	-5.36
3	5115	8.00	23652.0	27.18
4	3523	7.73	23029.4	18.63
5	3666	6.75	20942.7	7.34
Overall	39268		19634.5	

Test

Null hypothesis H_0: All medians are equal

Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	5	2107.63	0.000
Adjusted for ties	5	2112.40	0.000

RQ3. Does taxi time differ by the number of hot spots on different airport hub classifications? This question was answered individually for small, medium, and large hub airport data. One-way ANOVA, Welch's test, and Kruskal-Wallis test were used to for analyses.

<u>**Small Hub Airports**</u> (Refer to Table 3 for detailed statistics)

<u>One-way ANOVA:</u> Using the data collected, an alpha of 0.05, and one-way ANOVA (assuming equal variances) test to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across the number of hotspots on small hub airports. Note – there were no small hub airports with 3 hotspots in the data.

> *Taxi-out time:* Using the Tukey HSD post hoc and 95% confidence, two significantly different groups (A and B) of small hub airports emerged. The mean taxi-out time was significantly different between the small hub airports in group A (0, 2, or 5 hotspots) and group B (1 or 4 hotspots). There was no significant difference within groups.
>
> *Taxi-in time:* Using the Tukey HSD post hoc and 95% confidence, three significantly different groups (A, B, and C) of small hub airports emerged. The mean taxi-in time was significantly different among the small hub airports in group A (2 hotspots), group B (0, 4, 5 hotspots), and group C (1, 4, 5 hotspots). There was no significant difference within groups.

<u>Welch's Test:</u> Using the data collected, an alpha of 0.05, and one-way ANOVA (not assuming equal variances) test to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across the number of hotspots on small hub airports. Note – there were no small hub airports with 3 hotspots in the data.

> *Taxi-out time:* Using the Games-Howell post hoc and 95% confidence, two significantly different groups (A and B) of small hub airports emerged. The mean taxi-out time was significantly different between the small hub airports in group A (0, 2, or 5 hotspots) and group B (1 or 4 hotspots). There was no significant difference within groups
>
> *Taxi-in time:* Using the Games-Howell post hoc and 95% confidence, three significantly different groups (A, B, and C) of small hub airports emerged. The mean taxi-in time was significantly different among the small hub airports in group A (2 hotspots), group B (0, 4, 5 hotspots), and group C (1, 4, 5 hotspots). There was no significant difference within groups.

<u>Kruskal-Wallis Test:</u> Using the data collected, an alpha of 0.05, and the Kruskal-Wallis non-parametric test to compare medians, the researchers *rejected the null hypotheses* (p-value<0.001) that the median taxi-out (and taxi-in) time was same across the number of hotspots on small hub airports. Therefore, the researchers concluded that median taxi-out (and taxi-in) time was different for at least one hotspot count. The mean ranks and Z-value indicate that the taxi-out (and taxi-in) times at airports with 0, 2, and 5 hotspots tend to be higher than those at airports with 1 and 4 hotspots. Mean rank of airports with 0, 2, and 5 hotspots was greater than the overall mean rank, whereas, mean ranks of airports with 1 and 4 hotspots were less than the overall mean rank. Note – there were no small hub airports with 3 hotspots in the data.

Table 3. Results for RQ3 – Small Hub Airports

RQ3 — Small Hub — Taxi time vs Hotspots

Taxi-out Time

Means

Small_Hot Spot	N	Mean	StDev	95% CI
0	3809	11.101	9.142	(10.839, 11.363)
1	3030	8.973	7.464	(8.680, 9.267)
2	2303	10.556	7.509	(10.219, 10.893)
4	963	9.035	7.357	(8.514, 9.555)
5	1115	10.671	9.122	(10.187, 11.155)

Pooled StDev = 8.24048

Taxi-in Time

Means

Small_Hot Spot	N	Mean	StDev	95% CI
0	3809	3.8120	3.9257	(3.6938, 3.9301)
1	3030	3.3539	3.9265	(3.2214, 3.4864)
2	2303	4.5176	3.4803	(4.3656, 4.6695)
4	963	3.563	3.873	(3.328, 3.798)
5	1115	3.5976	2.6011	(3.3792, 3.8160)

Pooled StDev = 3.72044

One-way ANOVA (Assuming equal variances) — Post hoc: Tukey HSD

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times at *small* hubs)
H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Small_Hot Spot	4	9497	2374.34	34.97	0.000
Error	11215	761561	67.91		
Total	11219	771059			

Grouping Information Using the Tukey Method and 95% Confidence

Small_Hot Spot	N	Mean	Grouping
0	3809	11.101	A
5	1115	10.671	A
2	2303	10.556	A
4	963	9.035	B
1	3030	8.973	B

Means that do not share a letter are significantly different.

Taxi-in Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Small_Hot Spot	4	1888	472.00	34.10	0.000
Error	11215	155235	13.84		
Total	11219	157123			

Grouping Information Using the Tukey Method and 95% Confidence

Small_Hot Spot	N	Mean	Grouping	
2	2303	4.5176	A	
0	3809	3.8120	B	
5	1115	3.5976	B	C
4	963	3.563	B	C
1	3030	3.3539		C

Means that do not share a letter are significantly different.

Welch's Test (Not assuming equal variances) — Post hoc: Games-Howell

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times at *small* hubs)
H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Small_Hot Spot	4	3724.99	36.82	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

Small_Hot Spot	N	Mean	Grouping
0	3809	11.101	A
5	1115	10.671	A
2	2303	10.556	A
4	963	9.035	B
1	3030	8.973	B

Means that do not share a letter are significantly different.

Taxi-in Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Small_Hot Spot	4	3859.71	36.83	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

Small_Hot Spot	N	Mean	Grouping	
2	2303	4.5176	A	
0	3809	3.8120	B	
5	1115	3.5976	B	C
4	963	3.563	B	C
1	3030	3.3539		C

Means that do not share a letter are significantly different.

Kruskal-Wallis (non-parametric)

H_o: $\eta_0 = \eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5$ (for both taxi-in and taxi-out times at *small* hubs)
H_a: median taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Descriptive Statistics

Small_Hot Spot	N	Median	Mean Rank	Z-Value
0	3809	12	5975.6	8.56
1	3030	10	5076.1	-10.63
2	2303	12	5917.4	5.10
4	963	10	5105.5	-5.06
5	1115	11	5617.8	0.08
Overall	11220		5610.5	

Test

Null hypothesis H_0: All medians are equal
Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	4	174.93	0.000
Adjusted for ties	4	179.28	0.000

Taxi-in Time

Descriptive Statistics

Small_Hot Spot	N	Median	Mean Rank	Z-Value
0	3809	4.0	5612.8	0.05
1	3030	3.0	4966.3	-12.81
2	2303	4.5	6481.5	14.47
4	963	4.0	5508.3	-1.02
5	1115	4.0	5642.7	0.35
Overall	11220		5610.5	

Test

Null hypothesis H_0: All medians are equal
Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	4	287.44	0.000
Adjusted for ties	4	296.74	0.000

Medium Hub Airports (Refer to Table 4 for detailed statistics)

One-way ANOVA: Using the data collected, an alpha of 0.05, and one-way ANOVA (assuming equal variances) test to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across the number of hotspots on medium hub airports. Note – there were no medium hub airports with 4 hotspots in the data.

> *Taxi-out time:* Using the Tukey HSD post hoc and 95% confidence, three significantly different groups (A, B and C) of medium hub airports emerged. The mean taxi-out time was significantly different among the medium hub airports in group A (5 hotspots), group B (0 hotspots) and group C (1, 2, 3 hotspots). There was no significant difference within groups.
>
> *Taxi-in time:* Using the Tukey HSD post hoc and 95% confidence, four significantly different groups (A, B, C, and D) of medium hub airports emerged. The mean taxi-in time was significantly different among the medium hub airports in group A (0 hotspots), group B (5 hotspots), group C (2 hotspots), and group D (1 or 3 hotspots). There was no significant difference within groups.

Welch's Test: Using the data collected, an alpha of 0.05, and one-way ANOVA (not assuming equal variances) test to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across the number of hotspots on medium hub airports. Note – there were no medium hub airports with 4 hotspots in the data.

> *Taxi-out time:* Using the Games-Howell post hoc and 95% confidence, three significantly different groups (A, B and C) of medium hub airports emerged. The mean taxi-out time was significantly different among the medium hub airports in group A (5 hotspots), group B (0 hotspots) and group C (1, 2, 3 hotspots). There was no significant difference within groups.
>
> *Taxi-in time:* Using the Games-Howell post hoc and 95% confidence, four significantly different groups (A, B, C, and D) of medium hub airports emerged. The mean taxi-in time was significantly different among the medium hub airports in group A (0 hotspots), group B (5 hotspots), group C (2 hotspots), and group D (1 or 3 hotspots). There was no significant difference within groups.

Kruskal-Wallis Test: Using the data collected, an alpha of 0.05, and the Kruskal-Wallis non-parametric test to compare medians, the researchers *rejected the null hypotheses* (p-value<0.001) that the median taxi-out (and taxi-in) time was same across the number of hotspots on medium hub airports. Therefore, the researchers concluded that median taxi-out (and taxi-in) time was different for at least one hotspot count. The mean ranks and Z-value indicate that the taxi-out times at airports with 0 and 5 hotspots tend to be higher than those at airports with 1, 2, and 3 hotspots. For taxi-out times, the mean rank of airports with 0 and 5 hotspots was greater than the overall mean rank, whereas, mean ranks of airports with 1, 2 and 3 hotspots were less than the overall mean rank. The mean ranks and Z-value indicate that the taxi-in times at airports with 0, 2, and 5 hotspots tend to be higher than those at airports with 1 and 3 hotspots. For taxi-in times, the mean rank of airports with 0, 2, and 5 hotspots was greater than the overall mean rank, whereas, mean ranks of airports with 1 and 3 hotspots were less than the overall mean rank.

Table 4. Results for RQ3 – Medium Hub Airports

RQ3 – Medium Hub – Taxi time vs Hotspots

Taxi-out Time

Means

Medium_Hot Spot	N	Mean	StDev	95% CI
0	1263	13.245	6.399	(12.944, 13.545)
1	3835	12.2315	5.8083	(12.0591, 12.4038)
2	5044	12.0267	5.8085	(11.8765, 12.1770)
3	2557	12.3371	4.0400	(12.1261, 12.5482)
5	1271	14.515	4.082	(14.216, 14.815)

Pooled StDev = 5.44504

Taxi-in Time

Means

Medium_Hot Spot	N	Mean	StDev	95% CI
0	1263	8.253	4.385	(8.078, 8.428)
1	3835	5.1042	2.8860	(5.0039, 5.2045)
2	5044	5.9759	3.3227	(5.8884, 6.0633)
3	2557	5.2065	2.6742	(5.0837, 5.3294)
5	1271	6.6842	2.8154	(6.5099, 6.8584)

Pooled StDev = 3.16880

One-way ANOVA (Assuming equal variances) – Post hoc: Tukey HSD

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times at *medium* hubs)
H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Medium_Hot Spot	4	7328	1832.05	61.79	0.000
Error	13965	414041	29.65		
Total	13969	421369			

Grouping Information Using the Tukey Method and 95% Confidence

Medium_Hot Spot	N	Mean	Grouping		
5	1271	14.515	A		
0	1263	13.245		B	
3	2557	12.3371			C
1	3835	12.2315			C
2	5044	12.0267			C

Means that do not share a letter are significantly different.

Taxi-in Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Medium_Hot Spot	4	11447	2861.74	285.00	0.000
Error	13965	140227	10.04		
Total	13969	151674			

Grouping Information Using the Tukey Method and 95% Confidence

Medium_Hot Spot	N	Mean	Grouping			
0	1263	8.253	A			
5	1271	6.6842		B		
2	5044	5.9759			C	
3	2557	5.2065				D
1	3835	5.1042				D

Means that do not share a letter are significantly different.

Welch's Test (Not assuming equal variances) – Post hoc: Games-Howell

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times at *medium* hubs)
H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Medium_Hot Spot	4	4614.49	91.95	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

Medium_Hot Spot	N	Mean	Grouping		
5	1271	14.515	A		
0	1263	13.245		B	
3	2557	12.3371			C
1	3835	12.2315			C
2	5044	12.0267			C

Means that do not share a letter are significantly different.

Taxi-in Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Medium_Hot Spot	4	4441.90	217.07	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

Medium_Hot Spot	N	Mean	Grouping			
0	1263	8.253	A			
5	1271	6.6842		B		
2	5044	5.9759			C	
3	2557	5.2065				D
1	3835	5.1042				D

Means that do not share a letter are significantly different.

Kruskal-Wallis (non-parametric)

H_o: $\eta_0 = \eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5$ (for both taxi-in and taxi-out times at *medium* hubs)
H_a: median taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Descriptive Statistics

Medium_Hot Spot	N	Median	Mean Rank	Z-Value
0	1263	13.29	7923.6	8.67
1	3835	11.50	6082.1	-16.29
2	5044	12.40	6856.6	-2.84
3	2557	12.33	6901.4	-1.17
5	1271	14.67	9460.2	22.94
Overall	13970		6985.5	

Test

Null hypothesis H_0: All medians are equal
Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	4	745.60	0.000
Adjusted for ties	4	746.30	0.000

Taxi-in Time

Descriptive Statistics

Medium_Hot Spot	N	Median	Mean Rank	Z-Value
0	1263	7.80	9938.1	27.28
1	3835	4.50	5584.6	-25.26
2	5044	5.57	7326.0	7.50
3	2557	5.00	6067.4	-12.74
5	1271	6.25	8774.1	16.58
Overall	13970		6985.5	

Test

Null hypothesis H_0: All medians are equal
Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	4	1558.20	0.000
Adjusted for ties	4	1561.79	0.000

<u>Large Hub Airports</u> (Refer to Table 5 for detailed statistics)

<u>One-way ANOVA:</u> Using the data collected, an alpha of 0.05, and one-way ANOVA (assuming equal variances) to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across the number of hotspots on large hubs.

> *Taxi-out time:* Using the Tukey HSD post hoc and 95% confidence, five significantly different groups (A, B, C, D, and E) of large hub airports emerged. The mean taxi-out time was significantly different among the large hub airports in group A (0 hotspots), group B (4 hotspots), group C (3 hotspots), group D (1 hotspot), and group E (2 or 5 hotspots). There was no significant difference within groups.
>
> *Taxi-in time:* Using the Tukey HSD post hoc and 95% confidence, five significantly different groups (A, B, C, D, and E) of large hub airports emerged. The mean taxi-in time was significantly different among the large hub airports in group A (0 or 3 hotspots), group B (5 hotspots), group C (4 hotspots), group D (2 hotspot), and group E (1 hotspot). There was no significant difference within groups.

<u>Welch's Test:</u> Using the data collected, an alpha of 0.05, and one-way ANOVA (not assuming equal variances) to compare means, the researchers *rejected the null hypotheses* (p-value<0.001) that the mean taxi-out (and taxi-in) time was same across the number of hotspots on large hubs.

> *Taxi-out time:* Using the Games-Howell post hoc and 95% confidence, five significantly different groups (A, B, C, D, and E) of large hub airports emerged. The mean taxi-out time was significantly different among the large hub airports in group A (0 hotspots), group B (4 hotspots), group C (3 hotspots), group D (1 hotspot), and group E (2 or 5 hotspots). There was no significant difference within groups.
>
> *Taxi-in time:* Using the Games-Howell post hoc and 95% confidence, five significantly different groups (A, B, C, D, and E) of large hub airports emerged. The mean taxi-in time was significantly different among the large hub airports in group A (0 or 3 hotspots), group B (5 hotspots), group C (4 hotspots), group D (2 hotspot), and group E (1 hotspot). There was no significant difference within groups.

<u>Kruskal-Wallis Test:</u> Using the data collected, an alpha of 0.05, and the Kruskal-Wallis non-parametric test to compare medians, the researchers *rejected the null hypotheses* (p-value<0.001) that the median taxi-out (and taxi-in) time was same across the number of hotspots on large hub airports. Therefore, the researchers concluded that median taxi-out (and taxi-in) time was different for at least one hotspot count. The mean ranks and Z-value indicate that the taxi-out times at airports with 0, 3, and 4 hotspots tend to be higher than those at airports with 1, 2, and 5 hotspots. For taxi-out times, the mean rank of airports with 0, 3, and 4 hotspots was greater than the overall mean rank, whereas, mean ranks of airports with 1, 2 and 5 hotspots were less than the overall mean rank. The mean ranks and Z-value indicate that the taxi-in times at airports with 0, 3, and 5 hotspots tend to be higher than those at airports with 1, 2, and 4 hotspots. For taxi-in times, mean rank of airports with 0, 3, and 5 hotspots was greater than the overall mean rank, whereas, mean ranks of airports with 1, 2 and 4 hotspots were less than the overall rank.

Table 5. Results for RQ3 – Large Hub Airports

RQ3 — Large Hub — Taxi time vs Hotspots

Taxi-out Time

Means

Large_Hot SPot	N	Mean	StDev	95% CI
0	2560	22.232	8.349	(21.994, 22.470)
1	2560	18.565	7.210	(18.328, 18.803)
2	2560	17.1455	4.0907	(16.9078, 17.3832)
3	2558	19.388	5.755	(19.150, 19.626)
4	2560	19.953	5.507	(19.715, 20.190)
5	1280	17.0024	3.1997	(16.6663, 17.3385)

Pooled StDev = 6.13471

Taxi-in Time

Means

Large_Hot SPot	N	Mean	StDev	95% CI
0	2560	11.8181	4.4012	(11.6825, 11.9537)
1	2560	8.8322	2.7743	(8.6967, 8.9678)
2	2560	9.2403	2.7693	(9.1047, 9.3758)
3	2558	11.7511	3.3890	(11.6155, 11.8867)
4	2560	9.5728	4.1685	(9.4372, 9.7084)
5	1280	10.7106	2.7401	(10.5189, 10.9023)

Pooled StDev = 3.49946

One-way ANOVA (Assuming equal variances) — Post hoc: Tukey HSD

Ho: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times at *large* hubs)

H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Large_Hot SPot	5	43077	8615.33	228.92	0.000
Error	14072	529595	37.63		
Total	14077	572672			

Grouping Information Using the Tukey Method and 95% Confidence

Large_Hot SPot	N	Mean	Grouping
0	2560	22.232	A
4	2560	19.953	B
3	2558	19.388	C
1	2560	18.565	D
2	2560	17.1455	E
5	1280	17.0024	E

Means that do not share a letter are significantly different.

Taxi-in Time

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Large_Hot SPot	5	21243	4248.60	346.93	0.000
Error	14072	172328	12.25		
Total	14077	193571			

Grouping Information Using the Tukey Method and 95% Confidence

Large_Hot SPot	N	Mean	Grouping
0	2560	11.8181	A
3	2558	11.7511	A
5	1280	10.7106	B
4	2560	9.5728	C
2	2560	9.2403	D
1	2560	8.8322	E

Means that do not share a letter are significantly different.

Welch's Test (Not assuming equal variances) — Post hoc: Games-Howell

H_o: $\mu_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (for both taxi-in and taxi-out times at *large* hubs)

H_a: mean taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Large_Hot SPot	5	6240.30	260.76	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

Large_Hot SPot	N	Mean	Grouping
0	2560	22.232	A
4	2560	19.953	B
3	2558	19.388	C
1	2560	18.565	D
2	2560	17.1455	E
5	1280	17.0024	E

Means that do not share a letter are significantly different.

Taxi-in Time

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Large_Hot SPot	5	6028.55	375.31	0.000

Grouping Information Using the Games-Howell Method and 95% Confidence

Large_Hot SPot	N	Mean	Grouping
0	2560	11.8181	A
3	2558	11.7511	A
5	1280	10.7106	B
4	2560	9.5728	C
2	2560	9.2403	D
1	2560	8.8322	E

Means that do not share a letter are significantly different.

Kruskal-Wallis (non-parametric)

H_o: $\eta_0 = \eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5$ (for both taxi-in and taxi-out times at *large* hubs)

H_a: median taxi-out (or taxi-in) time is different for at least one hotspot count

Taxi-out Time

Descriptive Statistics

Large_Hot SPot	N	Median	Mean Rank	Z-Value
0	2560	19.405	8587.6	21.31
1	2560	16.475	6044.1	-13.70
2	2560	16.340	5577.4	-20.12
3	2558	19.000	7825.9	10.82
4	2560	18.780	7881.3	11.59
5	1280	16.625	5603.0	-13.26
Overall	14078		7039.5	

Test

Null hypothesis H_0: All medians are equal

Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	5	1221.90	0.000
Adjusted for ties	5	1221.92	0.000

Taxi-in Time

Descriptive Statistics

Large_Hot SPot	N	Median	Mean Rank	Z-Value
0	2560	11.00	8822.5	24.54
1	2560	8.29	5095.9	-26.75
2	2560	8.75	5774.6	-17.41
3	2558	11.33	9190.3	29.59
4	2560	8.67	5820.1	-16.78
5	1280	10.25	8030.9	9.15
Overall	14078		7039.5	

Test

Null hypothesis H_0: All medians are equal

Alternative hypothesis H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	5	2349.25	0.000
Adjusted for ties	5	2349.34	0.000

DISCUSSION

This paper analyzed taxi time data from a sample of 33 airports across three FAA hub classifications including Small, Medium, and Large. The researchers used the taxi time data as available in the ASPM dataset, which begs a series of questions regarding the ASPM dataset – what are criteria and measurements to maintain the data accuracy? The ASPM dataset gives definitions of taxi in and taxi out times, what are the authorities and references for such definition? How do the data collectors ensure the data accuracy and consistency when collecting data from 77 airports across the U.S. and across different airlines? What is the data publication process and how does the data publisher maintain consistent process for all airports? These questions are crucial to the analysis on airport taxi times. The ASPM dataset tracks, collects, and publishes operational data from 77 airports in the U.S., which are very different on both operational and geographic aspect. It is important to be consistent in both data collection and publication when dealing with different airports and airlines to minimize bias for the research that utilizes data from this dataset.

In this paper, taxi times are compared by number of hotspots (0, 1, 2, 3, 4, 5) on airports and across the three hub classifications. This research collected 11 samples from each airport categories because there were only 11 small hub airports reported in the NPIAS report; the same number of medium and large airports were collected to maintain equal sample size in each airport category. This resulted in no airports with 3 hotspots in the small hubs, and no airports with 4 hotspots in the medium hubs. The missing of hotspots in small and medium hubs could bias the analysis results or reduce the statistical power of this research.

This research paper may be used to teach statistical analyses and methodology when there is abundant real-world data which needs dirty data collection methods and complicated data cleaning and consolidation. The researchers demonstrated the use of parametric and non-parametric statistical techniques as one or more assumptions were not met. This is another learning lesson for students – how to proceed with tests and reach conclusions when the statistical assumptions fail but there is an abundance of data. In addition, many students are only familiar with junior level statistics that typically include parametric tests and not non-parametric tests. Therefore, this study can be used as an opportunity to explore non-parametric testing with real world data. Similarly, students may not learn the application and interpretation of two-way ANOVA in any junior statistics courses. Therefore, the researchers conducted rigorous tests using one-way ANOVA and t-tests, so that students may follow along. In future papers, more advanced statistical analysis tools such as multiple regression, simulation, and non-parametric analytics may be explored.

CONCLUSION

This research collected 33 sample airports across the three FAA hub classifications – small, medium, and large. Using parametric and non-parametric statistical tests, the researchers found that the mean (and median) taxi-out times and taxi-in times are different across the three hub classifications. Figure 3 indicates that as the airports get larger, the average taxi times tend to be larger. This result was aligned with the researchers' intuition.

Using parametric statistical analysis method, the researcher found that the mean taxi-out time was different for airports with 0, 3, 4, and 5 hotspots, and no difference was found in taxi-out time between airports with 1 and 2 hotspots. Similarly, the average taxi-in time was different for airports with 1, 2, 3, and 4 hotspots, and there were no significant differences found in airports with 0 and 5 hotspots. Using non-parametric methods, the researchers found that there were differences in median taxi-out and taxi-in time for at least one of the hotspot numbers (0, 1, 2, ,3, 4, 5).

The researchers also compared taxi times by number of hot spots on different airport hub classifications using parametric and non-parametric methods. 1) For small hub airports, the parametric methods suggest that the average taxi out time was different between the small hub airports in group A (0, 2, or 5 hotspots) and group B (1 or 4 hotspots); the average taxi in time was different among the small hub airports in group A (2 hotspots), group B (0, 4, 5 hotspots), and group C (1, 4, 5 hotspots). Using non-parametric tests, the researchers concluded that median taxi-out (and taxi-in) time was different for at least one hotspot count. 2) For medium hub airports, the researchers used parametric tests and found that the mean taxi-out time was different among the medium hub airports in group A (5 hotspots), group B (0 hotspots) and group C (1, 2, 3 hotspots); and the mean taxi-in time was different among the medium hub airports in group A (0 hotspots), group B (5 hotspots), group C (2 hotspots), and group D (1 or 3 hotspots). The non-parametric tests indicate that median taxi-out (and taxi-in) time was different for at least one hotspot count. 3). For large hub airports, the researchers found that the mean taxi-out time was different among the large hub airports in group A (0 hotspots), group B (4 hotspots), group C (3 hotspots), group D (1 hotspot), and group E(2 or 5 hotspots); and the mean taxi-in time was different among the large hub airports in group A (0 or 3 hotspots), group B (5 hotspots), group C (4 hotspots), group D (2 hotspot), and group E(1 hotspot). Using non-parametric methods, the researchers concluded that median taxi-out (and taxi-in) time was different for at least one hotspot count.

Future research will focus on comparing taxi-out and taxi-in times across the three NPIAS hub classifications and number of hotspots. By combining the results of this study and the future research, the researchers aim to better understand and model taxi-in and taxi-out times at small, medium, and large hub airports. In addition, the researchers will explore other potential variables such as weather conditions and runway configurations that may have a significant impact on taxi times at airports.

REFERENCES

[1] S. Ravizza, J. Chen, J. A. D. Atkin, E. K. Burke, and P. Stewart, "The trade-off between taxi time and fuel consumption in airport ground movement," *Public Transp.*, vol. 5, no. 1-2, pp. 25-40, 2013, doi: 10.1007/s12469-013-0060-1.

[2] T. Nikoleris, G. Gupta, and M. Kistler, "Detailed estimation of fuel consumption and emissions during aircraft taxi operations at Dallas/Fort Worth International Airport," *Transp. Res. D. Transp. Environ.*, vol. 16, no. 4, pp. 302-308, 2011, doi: 10.1016/j.trd.2011.01.007.

[3] Federal Aviation Administration, "Hot Spot Standardized Symbology," *Federal Aviation Administration*, Jan. 04, 2023 [Online]. Available: https://www.faa.gov/airports/runway_safety/hotspots [Accessed Feb. 12, 2023]

[4] Federal Aviation Administration, "Hot Spot Standardized Symbology," Federal Aviation Administration, Mar. 16, 2022 [Online]. Available: https://www.faa.gov/newsroom/hot-spot-standardized-symbology [Accessed Feb. 12, 2023]

[5] Federal Aviation Administration, "ASPM 77," *Federal Aviation Administration*, [Online]. Available: https://aspm.faa.gov/aspmhelp/index/ASPM_77.html [Accessed Feb. 05, 2023]

[6] Federal Aviation Administration, "National Plan of Integrated Airport Systems (NPIAS)," *Federal Aviation Administration*, Dec. 7, 2022. [Online]. Available: https://www.faa.gov/airports/planning_capacity/npias [Accessed: Feb. 05, 2023].

[7] G. Clare and A. G. Richards, "Optimization of Taxiway Routing and Runway Scheduling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1000-1013, Dec. 2011, doi: 10.1109/TITS.2011.2131650.

[8] S. Ravizza, J. A. D. Atkin, and E. K. Burke, "A more realistic approach for airport ground movement optimization with stand holding," *J. Sched.*, vol. 17, no. 5, pp. 507-520, 2013, doi: 10.1007/s10951-013-0323-3.

[9] Federal Aviation Administration, "Airport Categories," *Federal Aviation Administration*, Dec. 07, 2022 [Online]. Available: https://www.faa.gov/airports/planning_capacity/categories [Accessed: Feb. 05, 2023].

[10] Federal Aviation Administration, "National Plan of Integrated Airport Systems (NPIAS) 2023-2027: Appendix A – List of NPIAS Airports." *Federal Aviation Administration*, Dec. 6, 2022. [Online]. Available: https://www.faa.gov/airports/planning_capacity/npias/current/2023_NPIAS_Appendix_A [Accessed Feb. 05, 2023].

[11] Federal Aviation Administration, "Aviation System Performance Metrics (ASPM)," *Federal Aviation Administration*. [Online]. Available: https://aspm.faa.gov/aspmhelp/index/Aviation_System_Performance_Metrics_(ASPM).html#Definitions_of_Variables [Accessed Feb. 12, 2023]

[12] Federal Aviation Administration, "Aviation System Performance Metrics: Airport Analysis." *Federal Aviation Administration.* [Dataset]. Available: https://aspm.faa.gov/apm/sys/AnalysisAP.asp [Accessed Feb. 12, 2023]

[13] Federal Aviation Administration, "Aviation System Performance Metrics: Airport Taxi Times." *Federal Aviation Administration.* [Dataset]. Available: https://aspm.faa.gov/apm/sys/TaxiTimes.asp [Accessed Feb. 12, 2023]

[14] Federal Aviation Administration, "FAA Airport Diagrams." *Federal Aviation Administration*, May 17, 2022. [Dataset]. Available: https://www.faa.gov/airports/runway_safety/diagrams/ [Accessed Feb. 12, 2023]