US Air Transport Operation Analysis From 2009 to 2018

TEAM 5

Group Members

- Ting-Hsuan Chen
- Yipeng Guo
- Yihan Jia
- Kexi Pi
- Weijia Suo (leader)
- Zheming Xu

Table of Contents

- Project Description and Data Exploration
- Business Problems Regarding **Major Airlines**
- Business Problems Regarding **Major Airports**
- Summary and Inference Based on Analysis

1. Project Description and Data Exploration

Project Description:

The project is aimed to identify the most popular airlines and the busiest airports in the United States with their performance evaluation.

Project Dimensions: Airlines + Airports

Data Description: Three main tables

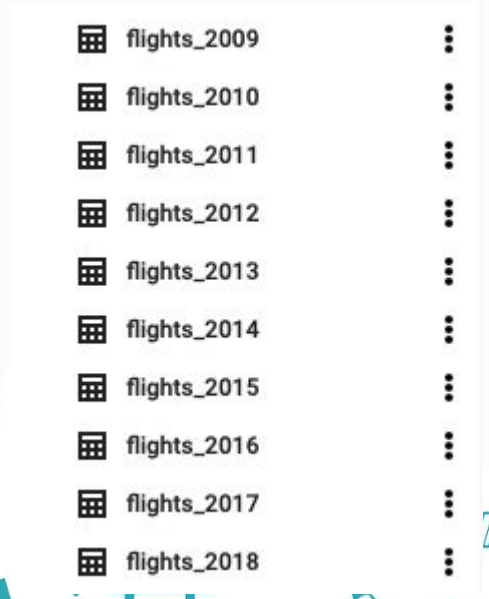
Flights Table: Store time-distinguished data for each flight around the U.S. from 2009 to 2018

Airports Table: Store all airport information in the United States

Airlines Table: Store all airline information for flight datasets

What We Have Done For Raw Data:

- **Step 1**: Renaming and casting each flights table after uploading the original dataset.
- **Step 2**: Combine all flights data and save to table flights_all



Field name	Type	Mode
FLIGHT_DATE	DATE	NULLABLE
IATA_CODE	STRING	NULLABLE
FLIGHT_NUMBER	INTEGER	NULLABLE
ORIGIN	STRING	NULLABLE
DESTINATION	STRING	NULLABLE
SCHEDULED_DEPARTURE	INTEGER	NULLABLE
DEPARTURE_TIME	INTEGER	NULLABLE
DEPARTURE_DELAY	INTEGER	NULLABLE
TAXI_OUT	INTEGER	NULLABLE
WHEELS_OFF	INTEGER	NULLABLE
WHEELS_ON	INTEGER	NULLABLE
TAXI_IN	INTEGER	NULLABLE
SCHEDULED_ARRIVAL	INTEGER	NULLABLE

- **Step 3**: Find all company name via IATA_CODE and store them in airlines table

```
%%bigquery --project ba775-team5
SELECT DISTINCT (IATA_CODE)
FROM `ba775-team5.us_airlines.flights_all`
LIMIT 10; -- Only for Preview
```

	IATA_CODE
0	OO
1	F9
2	OH
3	WN
4	YV
5	NW
6	XE
7	G4
8	CO
9	VX

airlines Table

IATA_CODE	STRING	NULLABLE
AIRLINE	STRING	NULLABLE

Data Reference: airlinecodes.info

- **Step 4**: Add time slot and year period attribute

Time slot: We divided the 24 hours into 12 groups. For example, 00;00 to 02:00 as group 1, 02:00 to 04:00 as group 2 etc.

Year Period: We divided the 10 years into 5 groups. For example, 2009 and 2010 as group 1 etc.

Data Preview

Time Range: Year 2009-2018

Important Schema:

Raw Data Source: Kaggle
airlinecodes.info

flight_all Table

Field name	Type	Mode	Policy Tags	Description
FLIGHT_DATE	DATE	NULLABLE		
IATA_CODE	STRING	NULLABLE		
FLIGHT_NUMBER	INTEGER	NULLABLE		
ORIGIN	STRING	NULLABLE		
DESTINATION	STRING	NULLABLE		
SCHEDULED_DEPARTURE	INTEGER	NULLABLE		
DEPARTURE_TIME	INTEGER	NULLABLE		
DEPARTURE_DELAY	INTEGER	NULLABLE		
TAXI_OUT	INTEGER	NULLABLE		
WHEELS_OFF	INTEGER	NULLABLE		
WHEELS_ON	INTEGER	NULLABLE		
TAXI_IN	INTEGER	NULLABLE		
SCHEDULED_ARRIVAL	INTEGER	NULLABLE		
ARRIVAL_TIME	INTEGER	NULLABLE		
ARRIVAL_DELAY	INTEGER	NULLABLE		
CANCELLED	INTEGER	NULLABLE		
CANCELLATION_CODE	STRING	NULLABLE		
DIVERTED	INTEGER	NULLABLE		
SCHEDULED_TIME	INTEGER	NULLABLE		
ELAPSED_TIME	INTEGER	NULLABLE		
AIR_TIME	INTEGER	NULLABLE		
DISTANCE	INTEGER	NULLABLE		
CARRIER_DELAY	INTEGER	NULLABLE		
WEATHER_DELAY	INTEGER	NULLABLE		
NAS_DELAY	INTEGER	NULLABLE		
SECURITY_DELAY	INTEGER	NULLABLE		
LATE_AIRCRAFT_DELAY	INTEGER	NULLABLE		
SCHEDULED_TIMESLOT	INTEGER	NULLABLE		
DEPARTURE_TIMESLOT	INTEGER	NULLABLE		
YEAR_PERIOD	INTEGER	NULLABLE		

Data Preview

Time Range: Year 2009-2018

Important Schema:

airlines Table

IATA_CODE	STRING	NULLABLE
AIRLINE	STRING	NULLABLE

airports Table

IATA_CODE	STRING	NULLABLE	Location Identifier
AIRPORT	STRING	NULLABLE	Airport's Name
CITY	STRING	NULLABLE	
STATE	STRING	NULLABLE	
COUNTRY	STRING	NULLABLE	Country Name of the Airport
LATITUDE	FLOAT	NULLABLE	Latitude of the Airport
LONGITUDE	FLOAT	NULLABLE	Longitude of the Airport

Raw Data Source: Kaggle

airlinecodes.info

Field name	Type	Mode	Policy Tags	Description

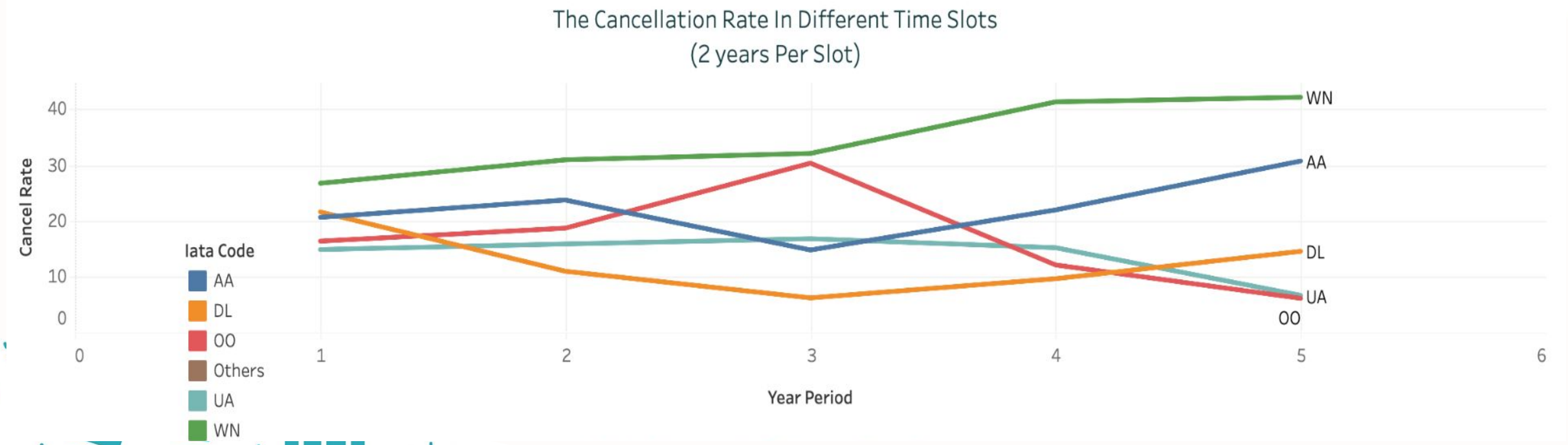
2. Business Problems Regarding *Major Airlines*

2.1 Top 5 US Airlines With The Most Flights In Past 10 Years

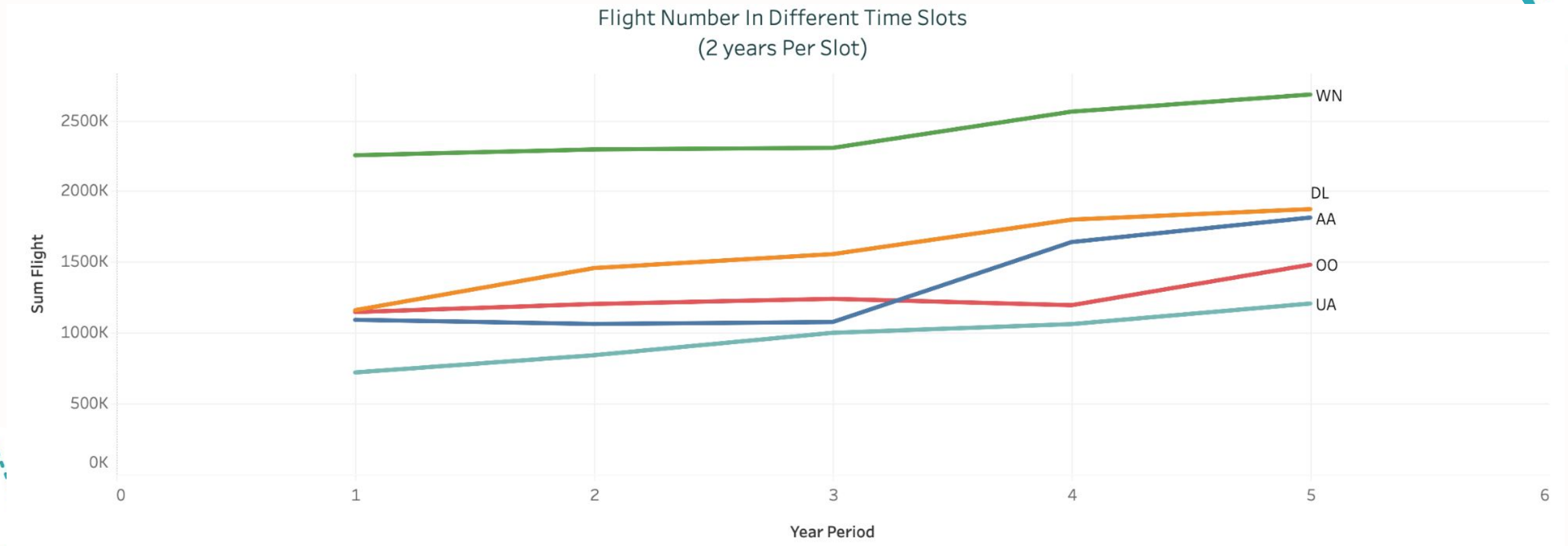
	IATA CODE	AIRLINE NAME	FLIGHT COUNT
1	WN	Southwest Airlines Co.	12,096,540
2	DL	Delta Air Lines Inc.	7,841,880
3	AA	American Airlines Inc.	6,682,161
4	OO	Skywest Airlines Inc.	6,263,052
5	UA	United Air Lines Inc.	4,826,658

2.2 The Cancellation Rate In Different Time Slots

(2 years Per Slot)



2.3 Flight Number In Different Time Slots
(2 years Per Slot)



2.4 Cancellation Rate Overall

(Past 10 years)

	IATA_CODE	Cancel_Rate
1.	WN	34.66
2.	UA	13.81
3.	OO	16.67
4.	DL	12.53
5.	AA	22.34

2.6 Top 3 Routes Of The Five Airlines

2.6.1 Delta Air Lines Inc.

	IATA CODE	AIRLINE NAME	FLIGHT COUNT	ORIGIN	DESTINATION
1	DL	Delta Air Lines Inc.	56,602	MCO	ATL
2	DL	Delta Air Lines Inc.	56,596	ATL	MCO
3	DL	Delta Air Lines Inc.	55,690	LGA	ATL

Atlanta: A transportation hub

2.6.2 American Airlines Inc.

	IATA CODE	AIRLINE NAME	FLIGHT COUNT	ORIGIN	DESTINATION
1	AA	American Airlines Inc.	57,504	DFW	LAX
2	AA	American Airlines Inc.	57,502	LAX	DFW
3	AA	American Airlines Inc.	55,994	DFW	ORD

Denver Fort Worth

2.6.3 United Air Lines Inc.

	IATA CODE	AIRLINE NAME	FLIGHT COUNT	ORIGIN	DESTINATION
1	UA	United Air Lines Inc.	47,840	SFO	LAX
2	UA	United Air Lines Inc.	47,422	LAX	SFO
3	UA	United Air Lines Inc.	45,793	SFO	ORD

San francisco International Airport
Los Angeles International Airport

2.6.4 Skywest Airlines Inc.

	IATA CODE	AIRLINE NAME	FLIGHT COUNT	ORIGIN	DESTINATION
1	OO	Skywest Airlines Inc.	56,110	SAN	LAX
2	OO	Skywest Airlines Inc.	56,100	LAX	SAN
3	OO	Skywest Airlines Inc.	33,827	MKE	ORD

San Diego International Airport
Los Angeles International Airport

2.6.5 Southwest Airlines Co.

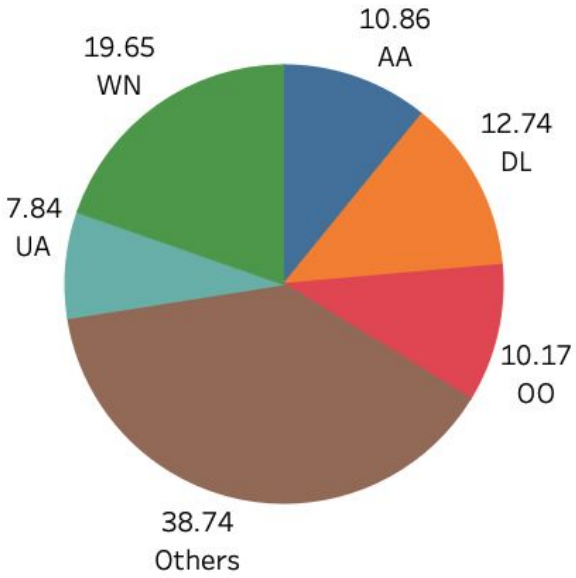
	IATA CODE	AIRLINE NAME	FLIGHT COUNT	ORIGIN	DESTINATION
1	WN	Southwest Airlines Co.	77,054	HOU	DAL
2	WN	Southwest Airlines Co.	76,957	DAL	HOU
3	WN	Southwest Airlines Co.	46,212	OAK	LAX

Dallas Love Field Airport

3. Business Problems Regarding *Major Airports*

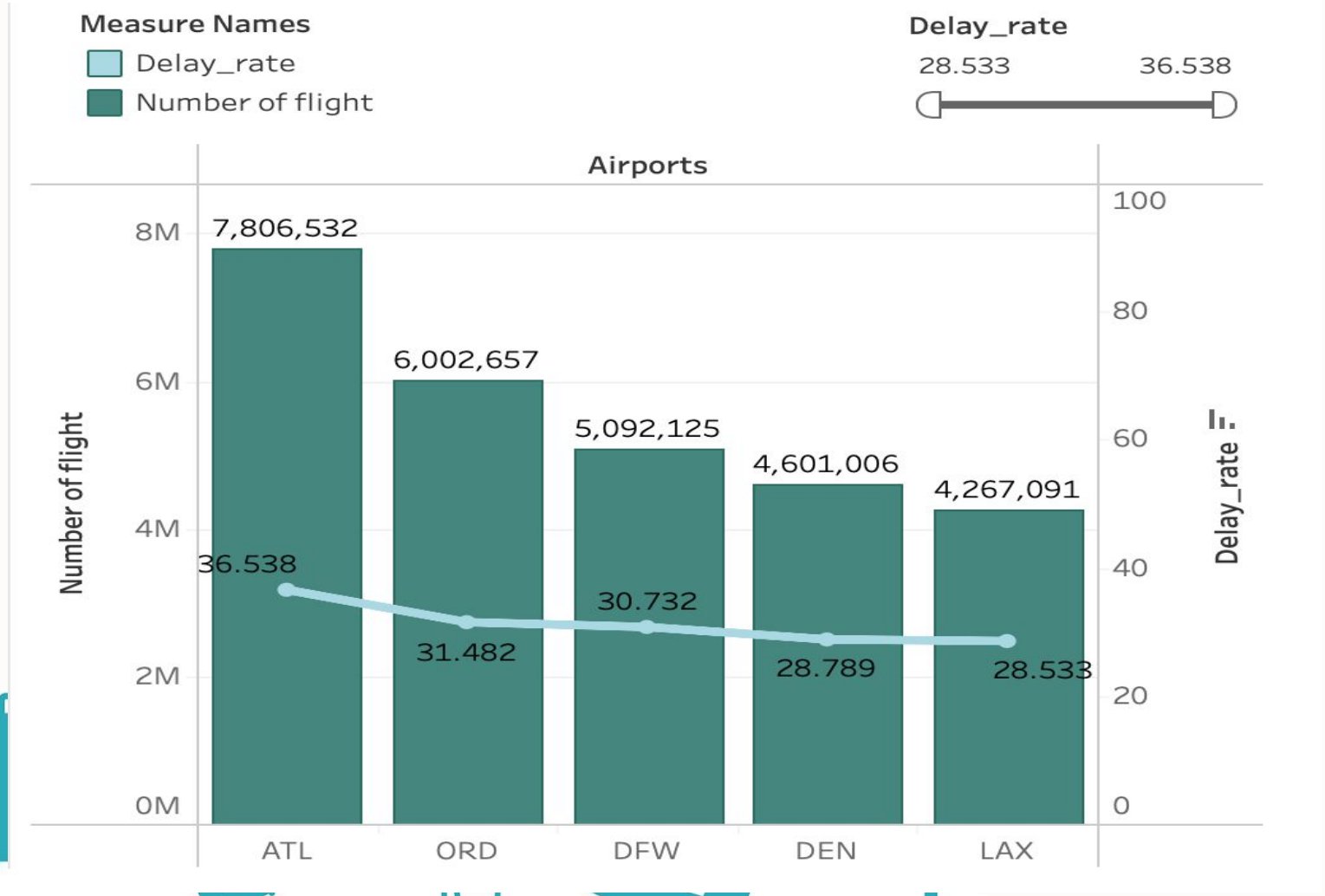
3.1 Top 5 Airports with the Most Flights: Last 10 years

	AIRPORTS	CITY	NUMBER OF FLIGHT
1	ATL	Atlanta	7,806,532
2	ORD	Chicago	6,002,657
3	DFW	Dallas-Fort Worth	5,092,125
4	DEN	Denver	4,601,006
5	LAX	Los Angeles	4,267,091



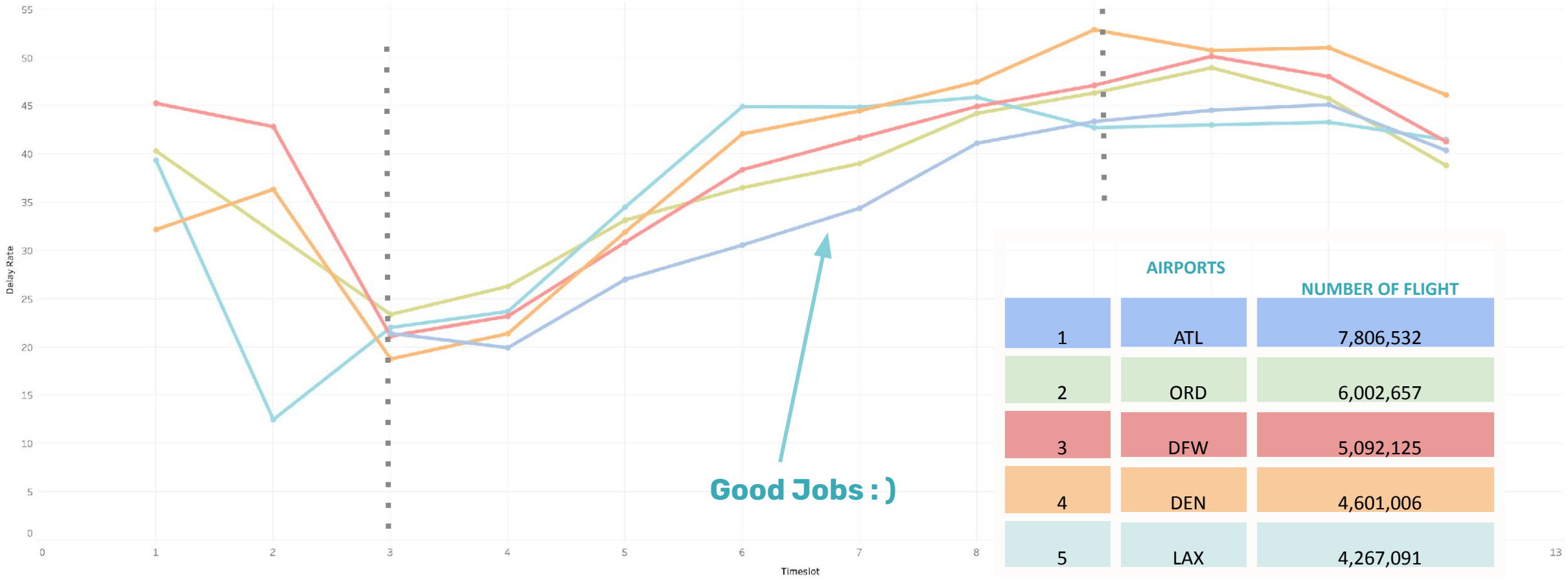
3.1.1 Geographical Distribution of Airports

3.2 Average Departure Delay Rate & Number of flight



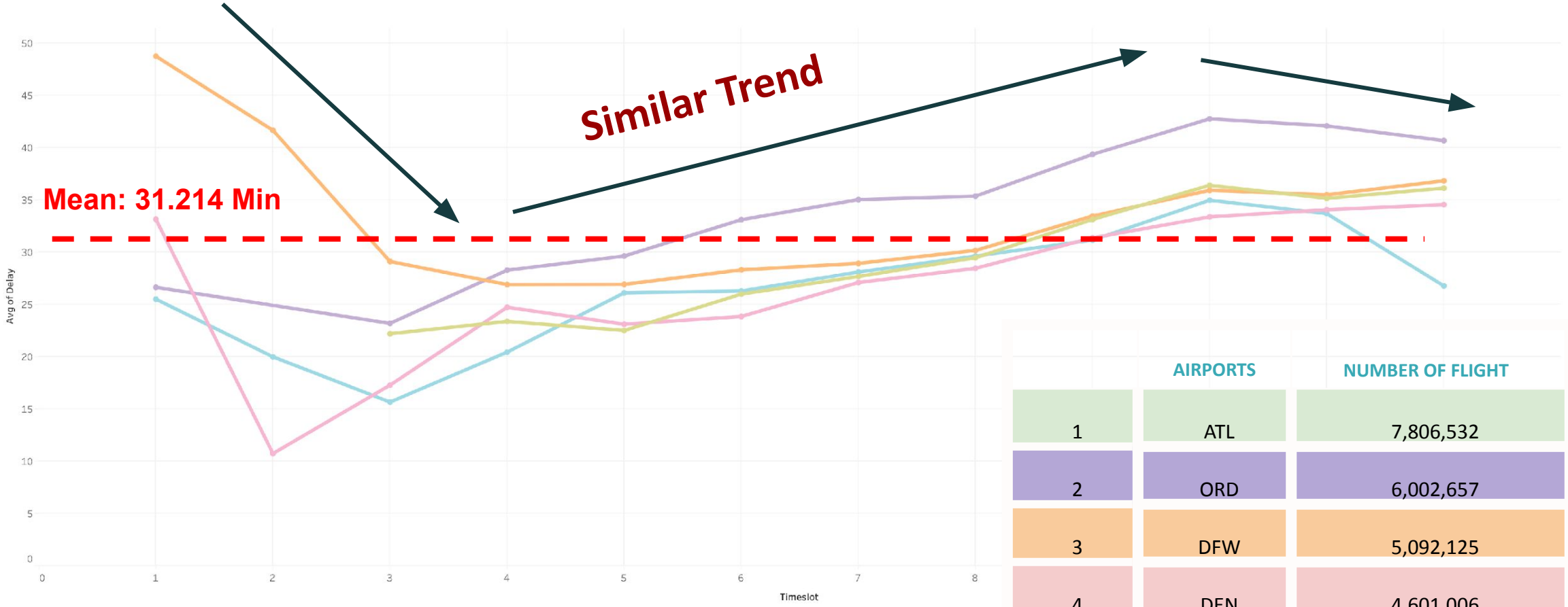
3.3.1 The Average Delay Rate (5 Airports)

Calculation Method: Delay Rate = Number_of_Delay / Number_of_flights



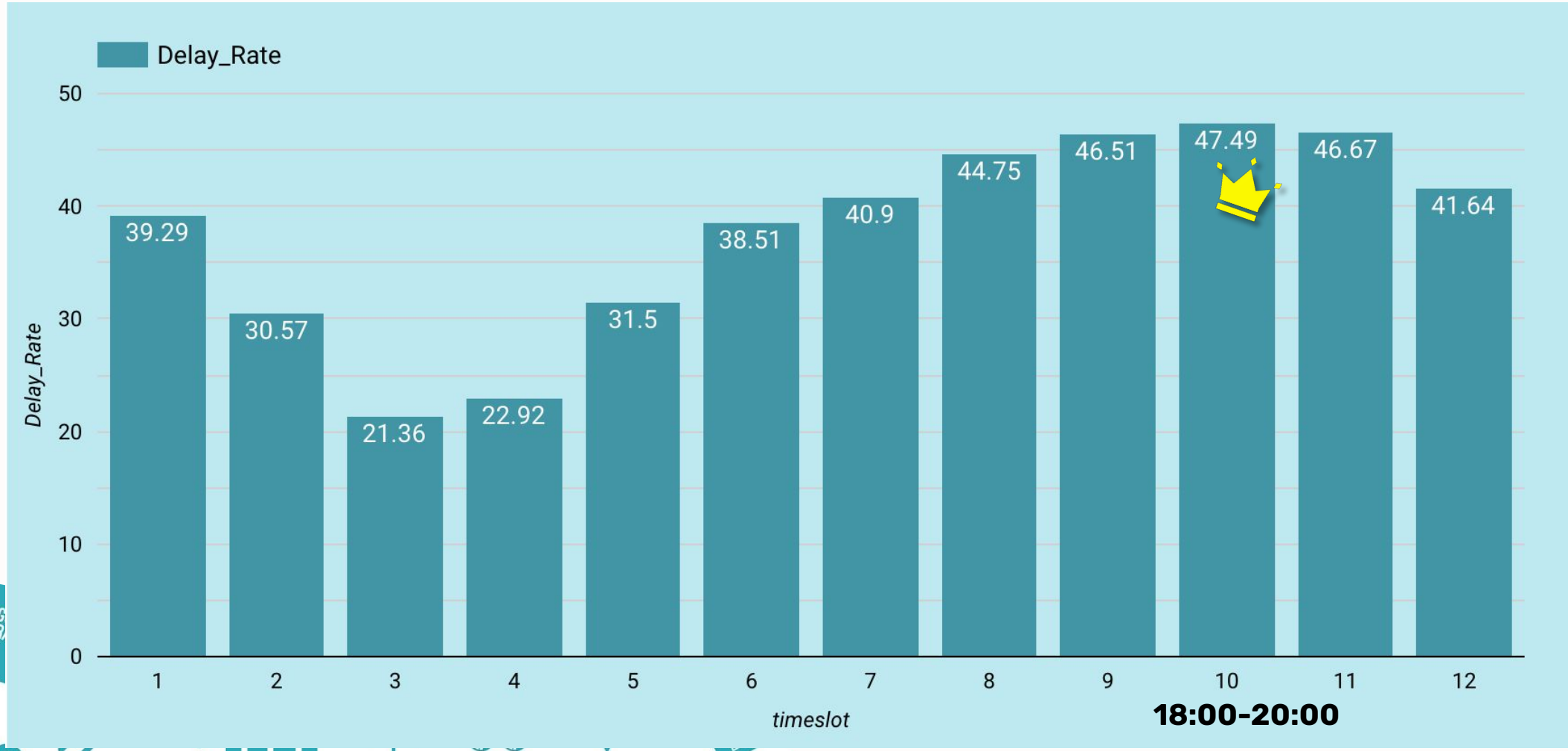
	AIRPORTS	NUMBER OF FLIGHT
1	ATL	7,806,532
2	ORD	6,002,657
3	DFW	5,092,125
4	DEN	4,601,006
5	LAX	4,267,091

3.3.2 The Departure Delay Time

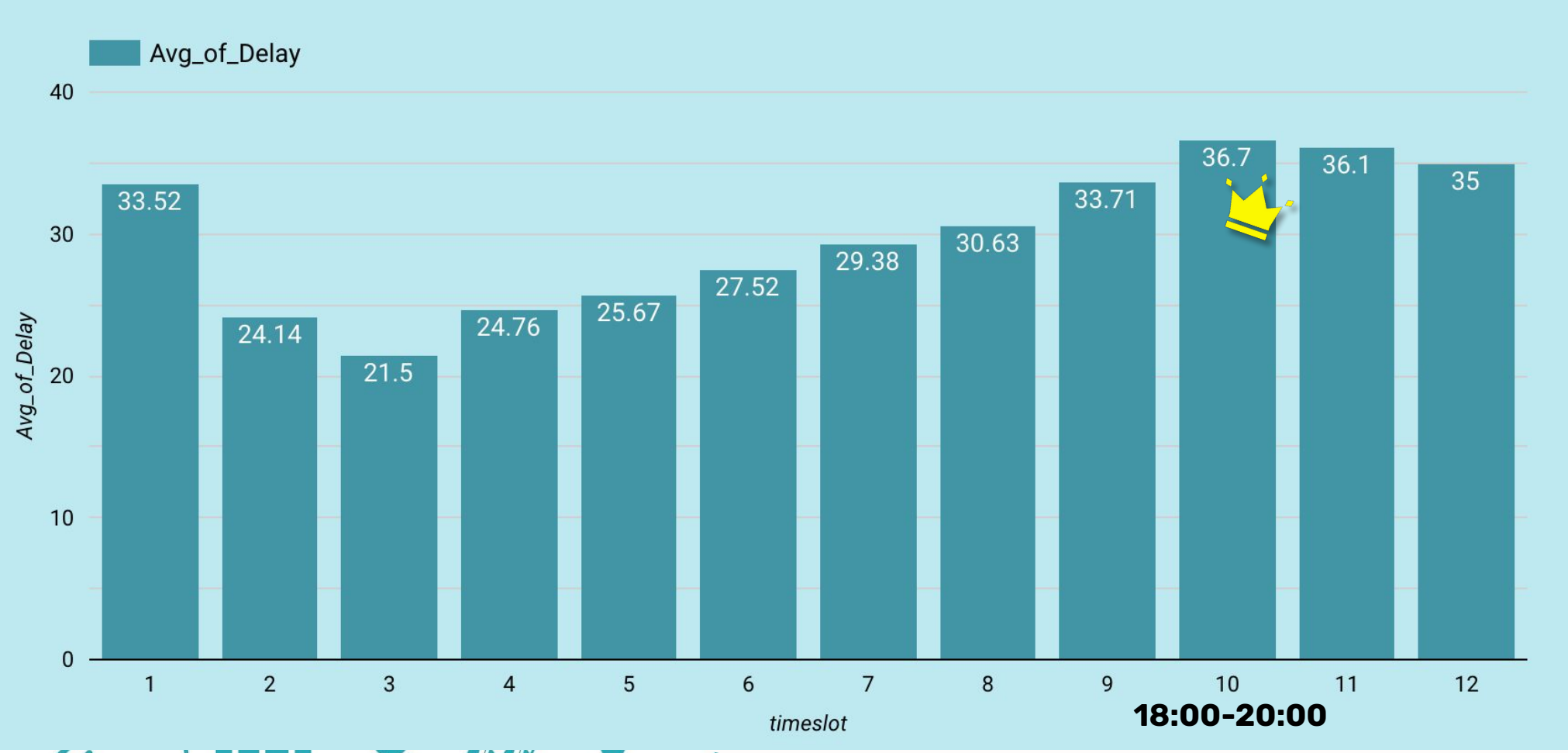


	AIRPORTS	NUMBER OF FLIGHT
1	ATL	7,806,532
2	ORD	6,002,657
3	DFW	5,092,125
4	DEN	4,601,006
5	LAX	4,267,091

3.3.3 The Average Delay Rate (Overall)

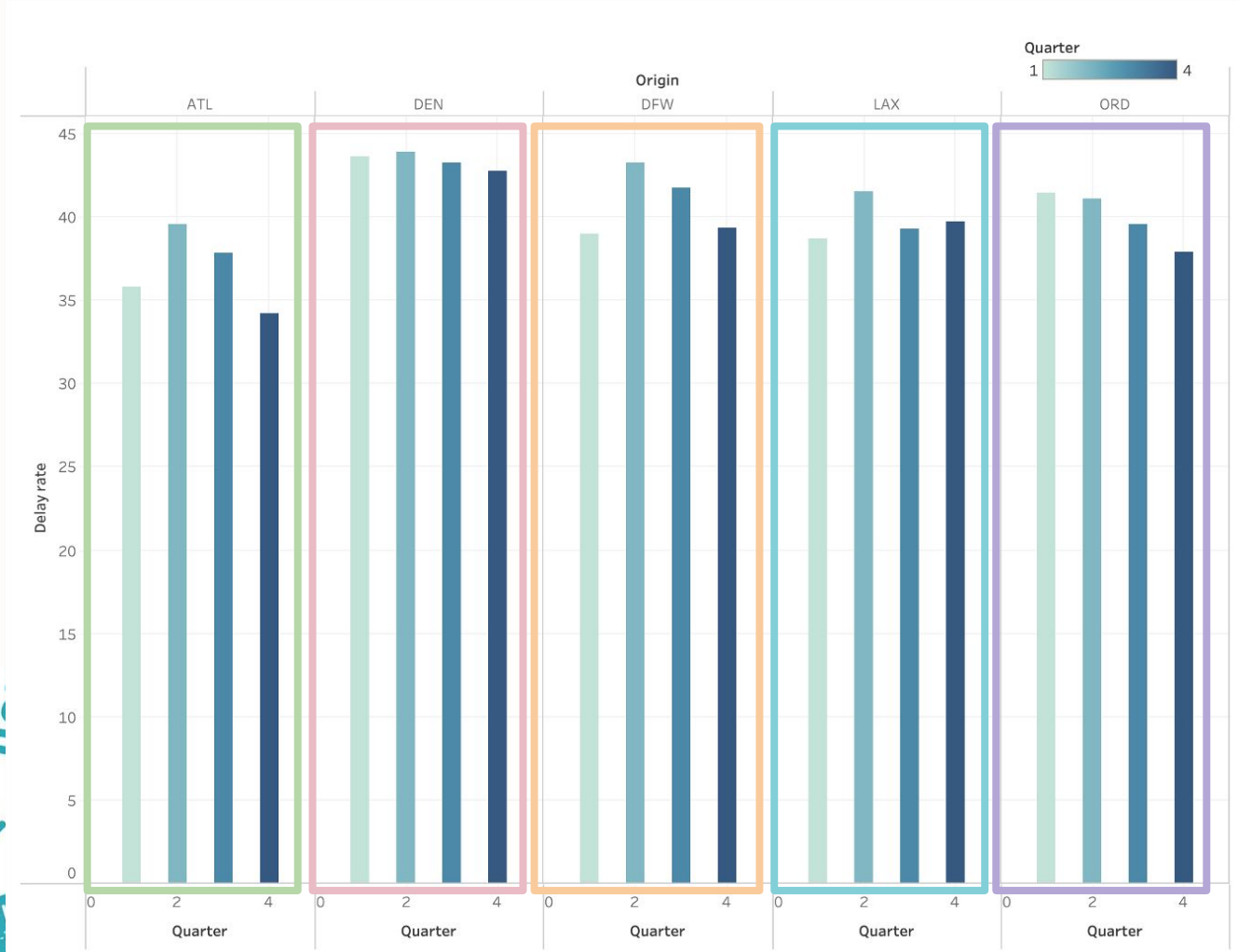


3.3.4 The Average Delay Time (Overall)



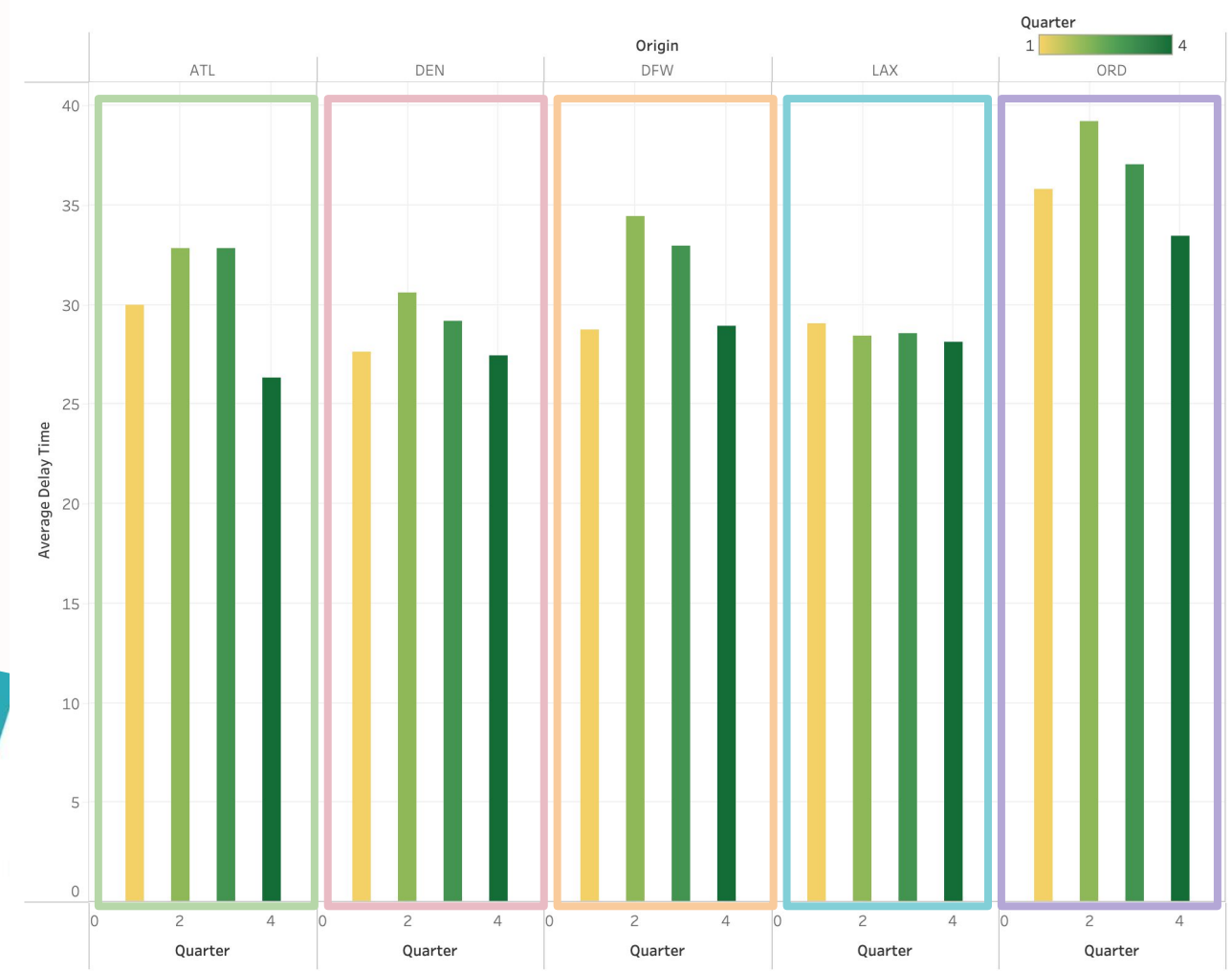
3.4 Dimension 2 - Quarter

3.4.1 The Delay Rate Per Quarter



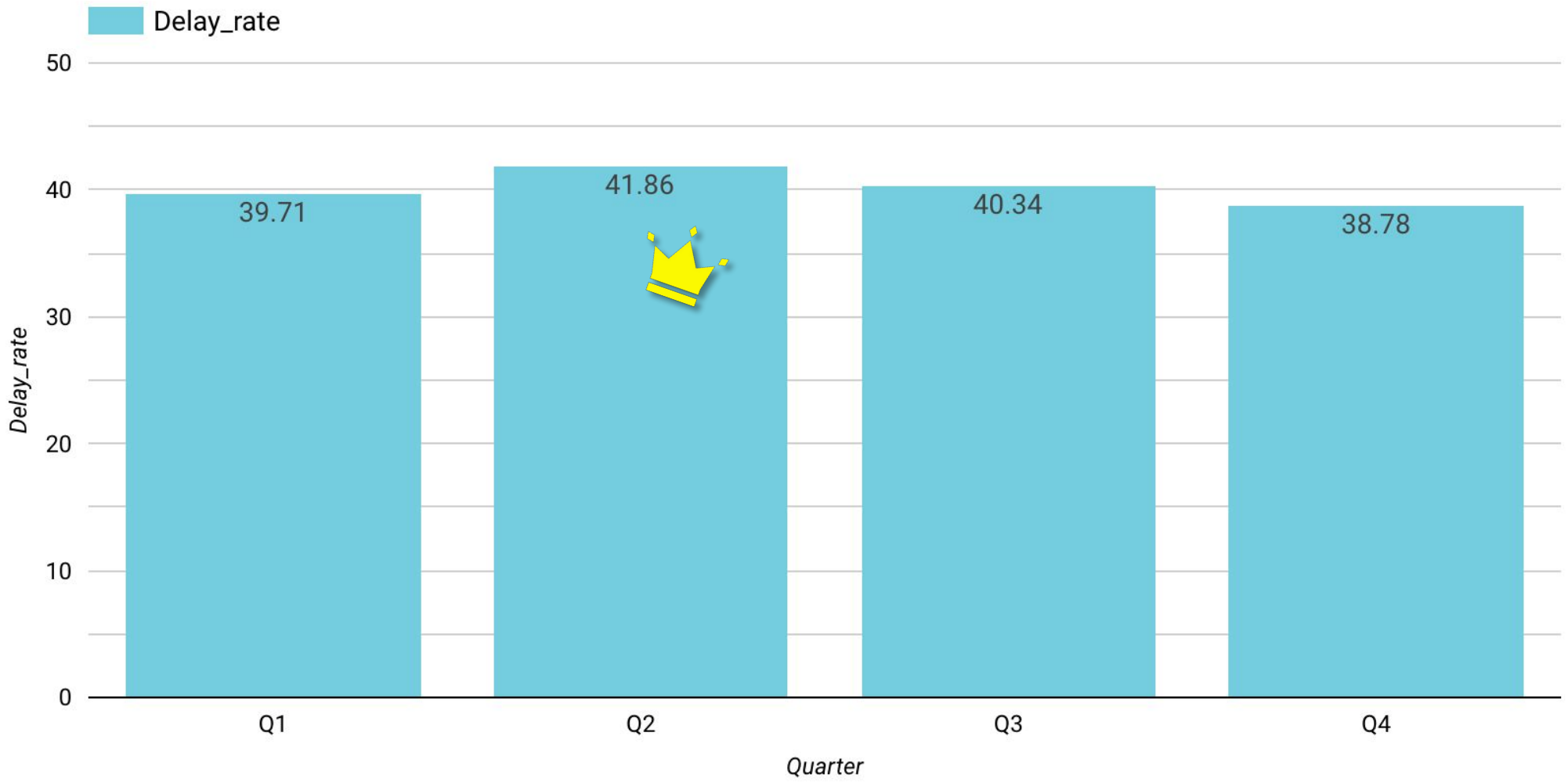
	AIRPORTS	NUMBER OF FLIGHT
1	ATL	7,806,532
2	ORD	6,002,657
3	DFW	5,092,125
4	DEN	4,601,006
5	LAX	4,267,091

3.4.2 The Average Delay Time (Per Quarter) For The Five Airports

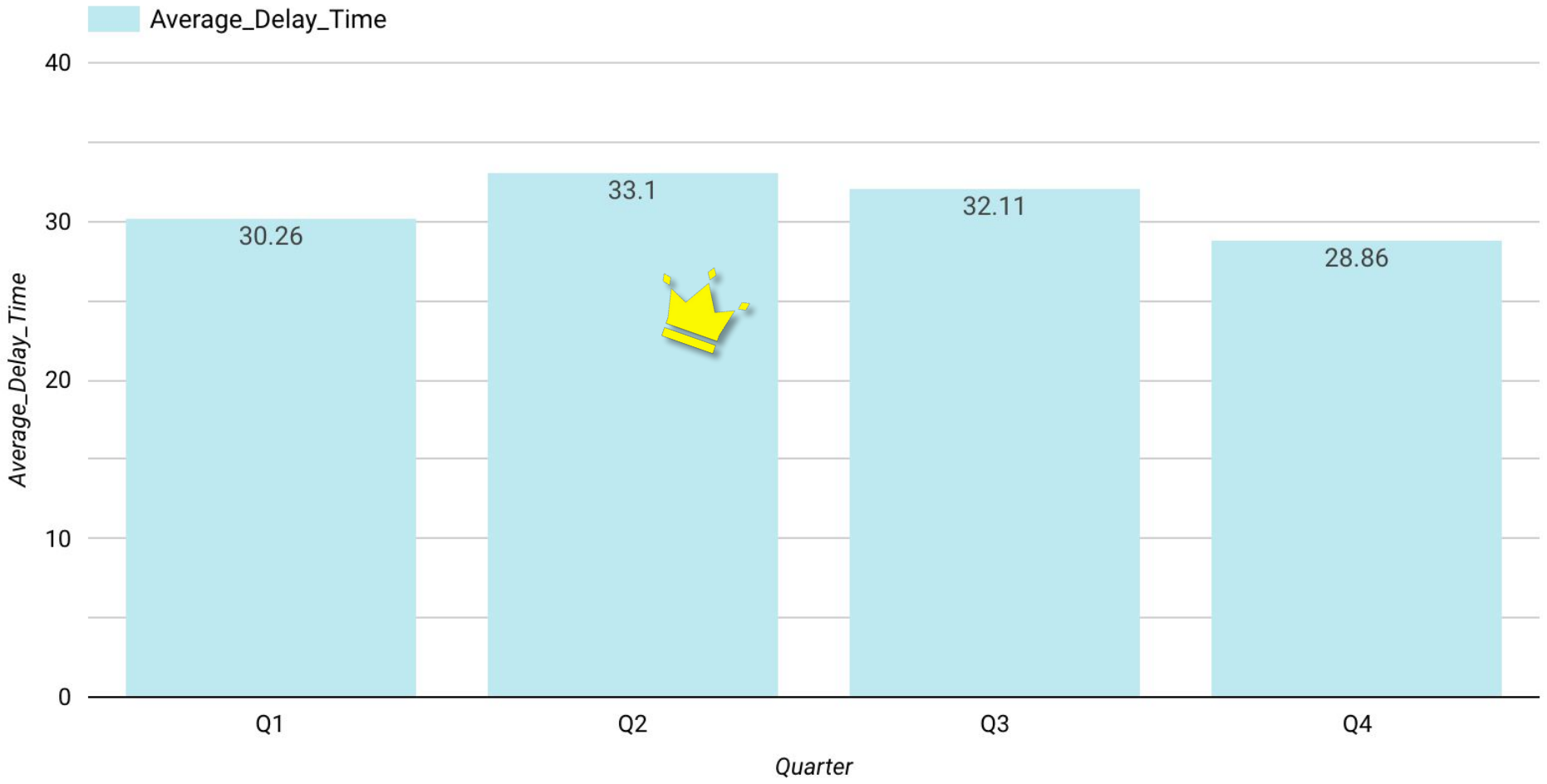


	AIRPORTS	NUMBER OF FLIGHT
1	ATL	7,806,532
2	ORD	6,002,657
3	DFW	5,092,125
4	DEN	4,601,006
5	LAX	4,267,091

3.4.3 The Average Delay Rate (Overall)



3.4.4 The Average Delay Time (Overall)



4. Summary and Inferences Based on Analysis

Top 5 Popular Airlines & Top 5 Busiest Airports

- **Airlines**: Delta Airlines(DL), American Airlines(AA), United Airlines(UA), Skywest Airlines(OO) and Southwest Airlines(WN).

- **Airports**: Atlanta(ATL), Chicago(ORD), Dallas-Fort Worth(DFW), Denver(DEN) and Los Angeles(LAX).


Washington
Montana
North Dakota
Minnesota
Maine
Oregon
Idaho
South Dakota
Wisconsin
VT
Wyoming
Michigan
New York
MA
Iowa
ORD
Chicago
United States
Ohio
NJ
Nevada
Utah
MD
West Virginia
DE
DEN
Denver
Missouri
Kentucky
California
Tennessee
North Carolina
Oklahoma
Arkansas
Arizona
New Mexico
Mississippi
LAX
Los Angeles
ATL
Atlanta
DFW
Dallas-Fort Worth
Florida

Analysis Towards the Airline Section

- Excluding objective factors such as weather, the **main reason for flight delays is that the number of flights is overloaded**, and the overall arrangement cannot be well arranged.

Analysis Towards the Airports Section

- In terms of overall quantity, we can find that **Atlanta Airport** has the largest number of flights and the lowest delay ratio, which shows that its operation is in good condition. **It is nice!**

- At the same time, from 2016 to 2018, Denver Airport ranked fourth in number of flights, but the delay rate was 5% higher than that of other airlines. **Operation issue here.**

Analysis Towards the Airports Section

- From the perspective of different time slots of a single day, slot_9, which is **8 to 10 p.m.** experiences the highest possibility of delay and cancel. **Passengers may avoid :)**

Analysis Towards the Airports Section

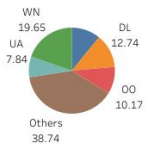
- From a quarterly point of view, the delay rate of **Denver Airport** and the delay time of **Los Angeles Airport** are relatively average throughout the four seasons. We can infer that it still has a relatively high delay rate in the off-season that this airport is in.
- Therefore, these two airports have **low seasons‘ operational issues**.

- On the whole, **Quarter 2 has the highest delay rate**, which may be related to people's travel preferences.

US Air Transport Operation Analysis From 2009 to 2018

Business Problems Regarding Major Airlines

The proportion of the five airlines' routes



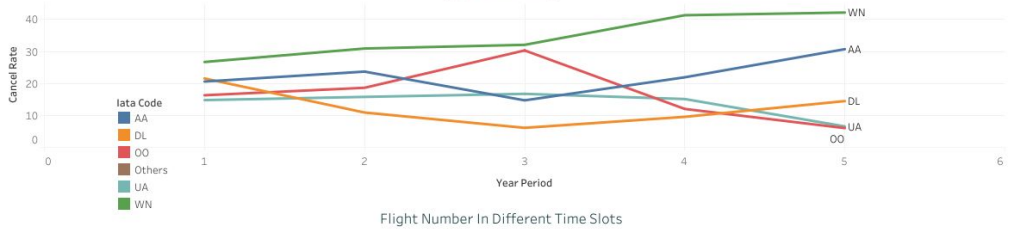
Top 5 US Airlines With The Most Flights Last 10 Years

Airline Name	Iata Code	
Southwest Airlines Co.	WN	12,096,540
Delta Air Lines Inc.	DL	7,841,880
American Airlines Inc.	AA	6,682,161
Skywest Airlines Inc.	OO	6,263,052
United Air Lines Inc.	UA	4,826,658

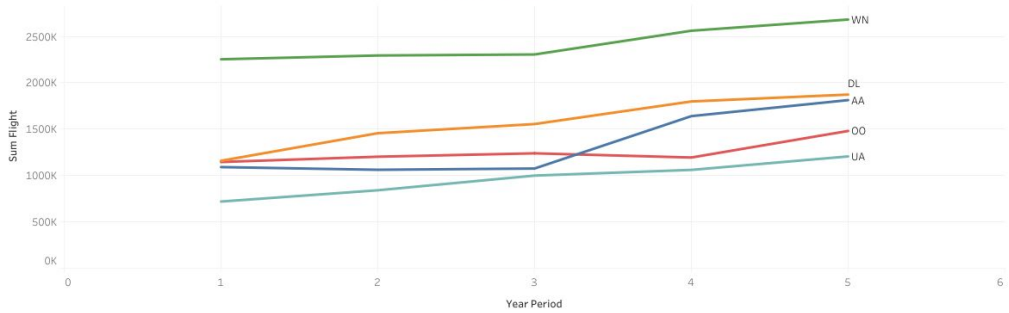
0M 2M 4M 6M 8M 10M 12M 14M

Flight Count

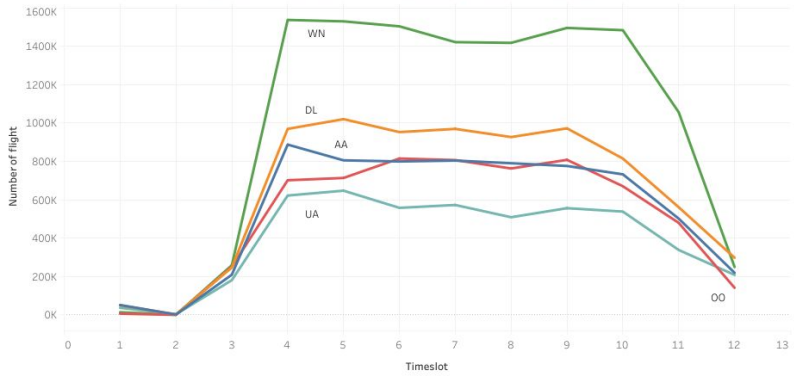
The Cancellation Rate In Different Time Slots
(2 years Per Slot)



Flight Number In Different Time Slots
(2 years Per Slot)



Flight Number Of The Five Airlines During Different Time Slots



Data Source

flights
Store time specific data for each flight from 2009 to 2018, such as departure time, arrival time, flight time, etc. Data comes from Kaggle. The original data size is 61,556,964 rows and about 7GB, the data is stored separately by year. For example, flights_2009, flight_2010.

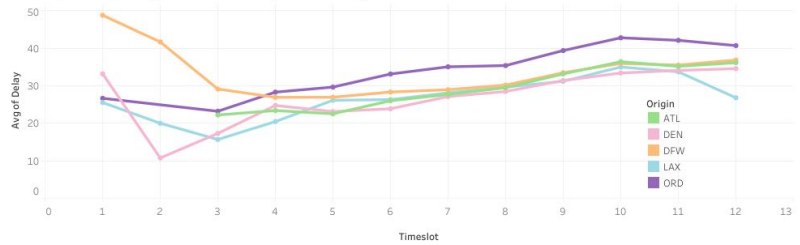
airports
Store all airport information in the United States, such as geographic information, airport code, etc. Data comes from Kaggle.

airlines
Store all airlines information for flights datasets, such as geographic information, airport code, etc. Table are created by our team and data are comes from airlinecodes.info

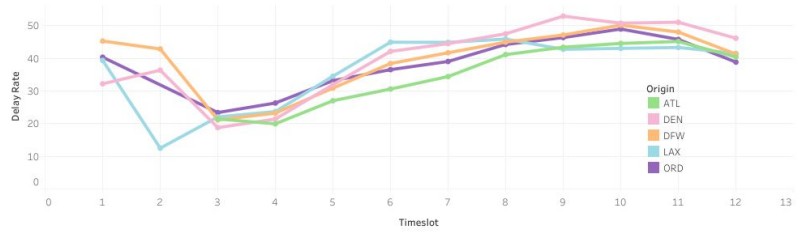
US Air Transport Operation Analysis From 2009 to 2018

Business Problems Regarding Major Airports

The Departure_Delay Of The Five Airports During Different Time Slots : 2hrs Per Slot

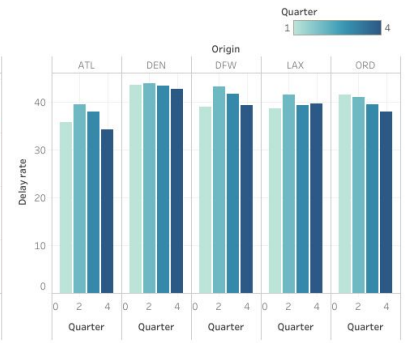


The Average Delay_Rate Of The Five Airports During Different Time Slots : 2hrs Per Slot

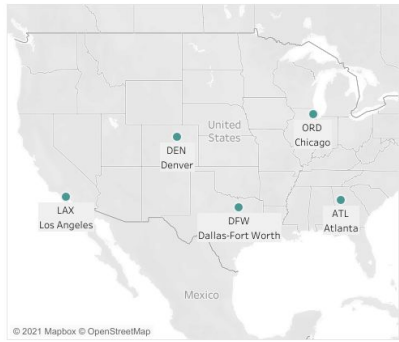


The Average Delay Time (Per Quarter) For The Five Airports

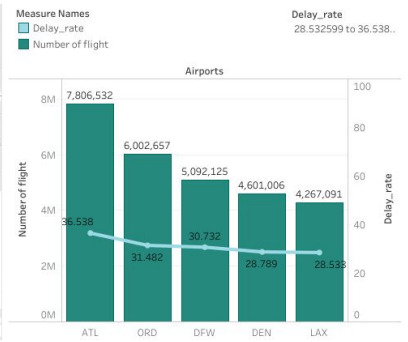
The Delay Rate Per Quarter For The Five Airports

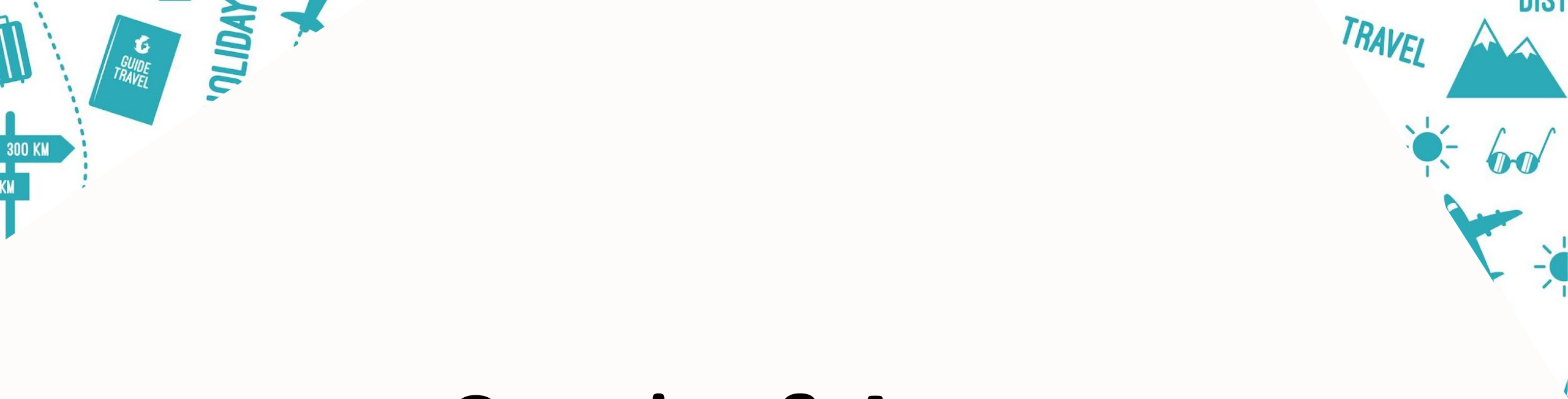


Geographical distribution of airports



Average Departure Delay Rate





Question & Answer

Thank You!