Principles of Data Science

Syed Jawad Hussain Shah

Flight Data Analysis

Team #Crusaders

Vanitha Kunta
Abhigna Reddy Mareddy
Krishna Saketh Channa
Manish Kumar Brungi

Introduction

Airline Data Analysis

- Used to identify patterns and trends in flight performance, safety, and operational efficiency.
- Help airlines optimize their flight operations, improve safety, and enhance the overall customer experience.
- The data gathered can help identify areas for improvement and inform decisions related to route planning, aircraft selection, and operational procedures.
- Through flight analysis, various metrics such as flight punctuality, flight duration, flight delays, and flight cancellations are monitored and evaluated.

Dataset

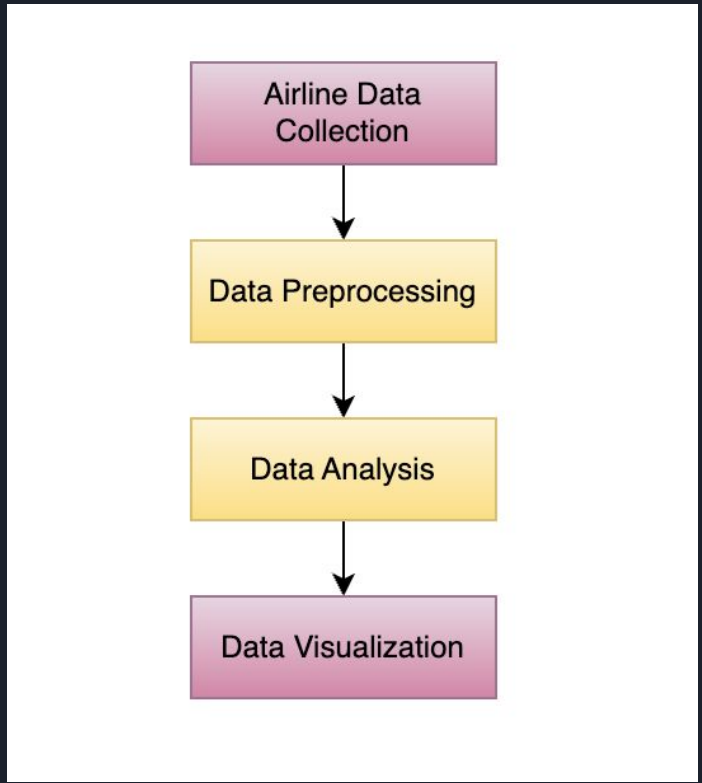
- The dataset is an Flight status Prediction dataset and has been chosen from Kaggle. The dataset has 61 columns and a few million rows.
- The dataset contains all flight information about cancellation and delays by different airlines.
- The dataset was downloaded from Kaggle for the years 2018, 2019, 2020, 2021 & 2022.
- https://www.kaggle.com/datasets/robikscube/flight-delay-dataset-20182022?select=Combined_Flights_2019.csv

Tools Used

- Framework
  - Flask
  - PySpark
- IDE
  - PyCharm
- Programming Language
  - Python
- Front-end
  - HTML, CSS, BootStrap, jQuery

Architecture

- Airline Data Collection - Kaggle DataSets based on years
- Data Preprocessing - Cleaning, Integration, Transformation etc.,
- Data Analysis
- Data visualization & User Interface



Data Preprocessing

- Collected all parquet datasets for efficient data storage and processing
- Combined all the 5 years datasets from 2018 - 2022
- **Preprocessing**:
  - Removed unnecessary columns
  - Extracted new columns from existing columns
  - **Ex**: From FlightDate (yyyy-MM-dd) column, extracted:
    - Year
    - Month
    - Day of Week
  - Added new columns

Data Preprocessing

- Extracted the column **DelayGroup** by categorizing the column **DepDelayMinutes** and with column **Cancelled** which is either True or False
  - **DepDelayMinutes** <= 0 : *OnTime_Early*
  - **DepDelayMinutes** <= 15 : *Small_Delay*
  - **DepDelayMinutes** <= 60 : *Medium_Delay*
  - **DepDelayMinutes** > 60 : *Large_Delay*
  - **Cancelled** == True : *Cancelled*
- Dataset size after preprocessing:

```
(29193782, 30)
```

Results

1. UI to visualize the airlines
2. Scheduled Flights
    a. Per Year
    b. By Airline
    c. By year for each Airline
3. Flight Delays
    a. By Year,Month,Week,Airline
    b. Delay Groups
    c. By Origin Airport and State
4. Distance travelled by Flights
5. Airports
    a. No. of flights On-time
    b. No. of flights Cancelled

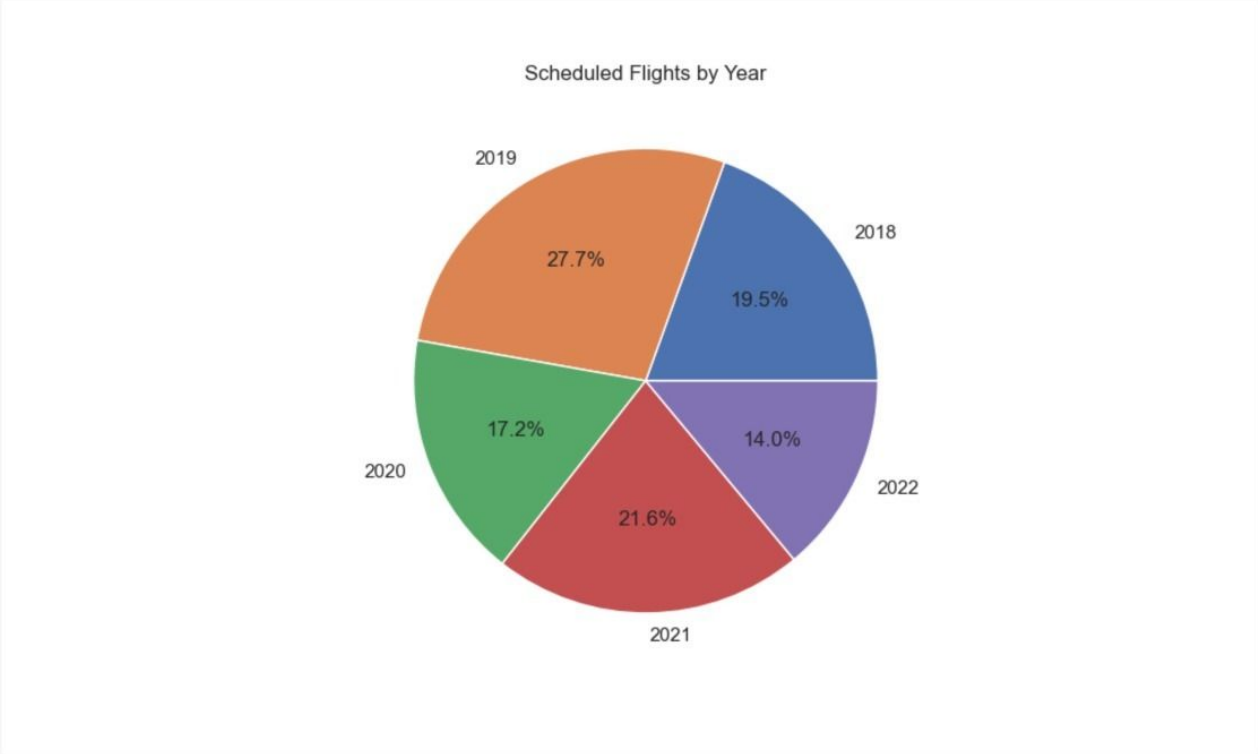
Show 25 entries

Search:

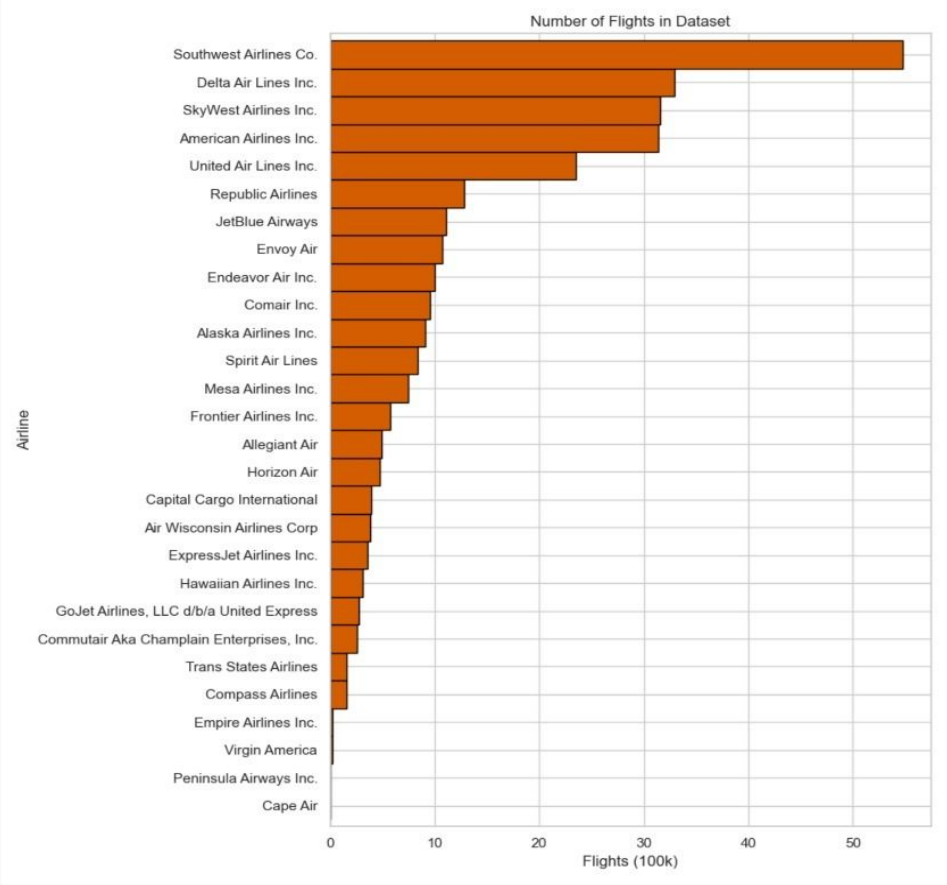
Airline Code	Name of the Airline
9E	Endeavor Air Inc.
9K	Cape Air
AA	American Airlines Inc.
AS	Alaska Airlines Inc.
AX	Trans States Airlines

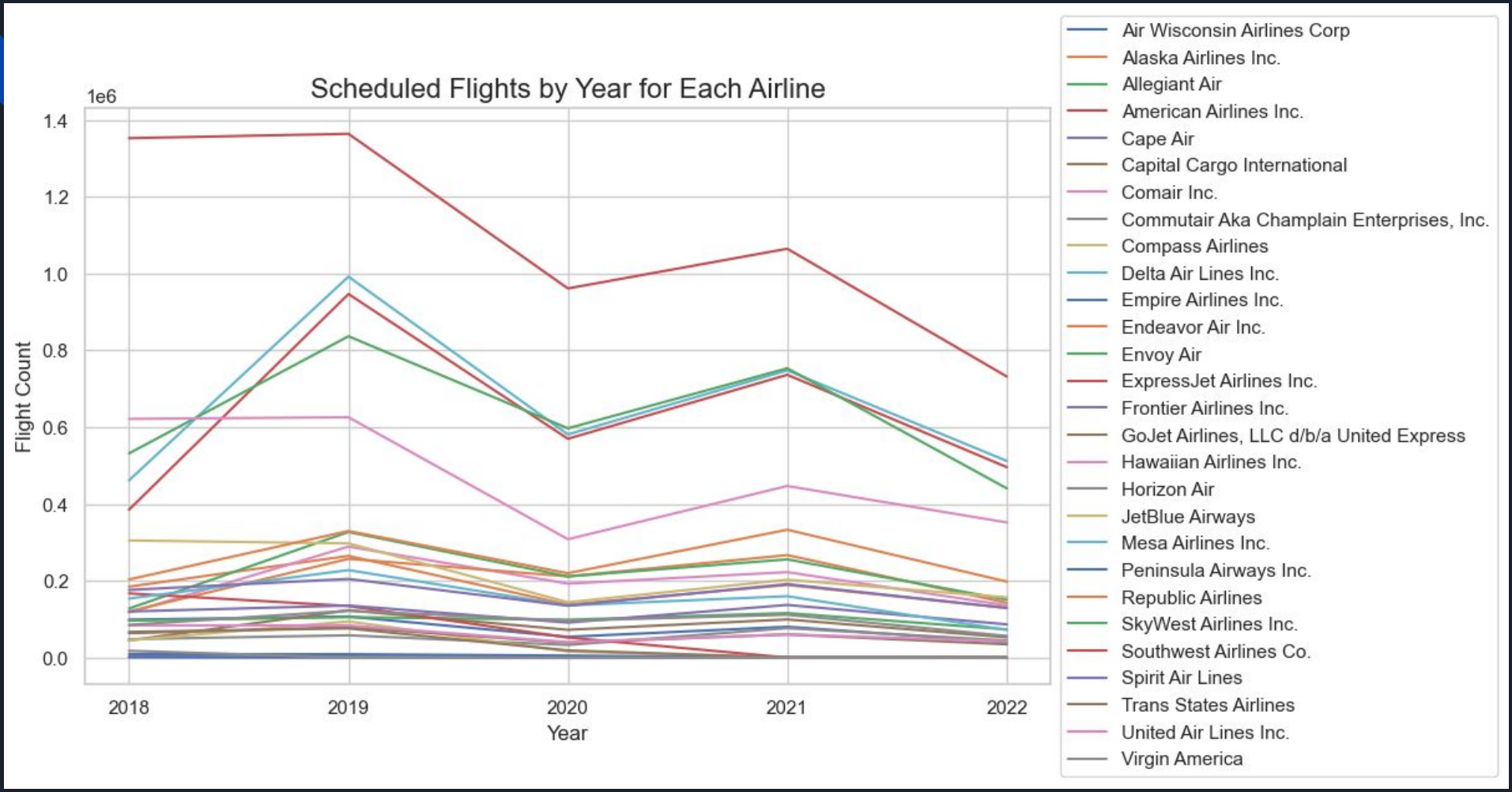
AX	Trans States Airlines
B6	JetBlue Airways
C5	Commutair Aka Champlain Enterprises, Inc.
CP	Compass Airlines
DL	Delta Air Lines Inc.
EM	Empire Airlines Inc.
EV	ExpressJet Airlines Inc.
F9	Frontier Airlines Inc.
G4	Allegiant Air
G7	GoJet Airlines, LLC d/b/a United Express
HA	Hawaiian Airlines Inc.
KS	Peninsula Airways Inc.
MQ	Envoy Air
NK	Spirit Air Lines
OH	Comair Inc.
OO	SkyWest Airlines Inc.
PT	Capital Cargo International
QX	Horizon Air
UA	United Air Lines Inc.
VX	Virgin America
WN	Southwest Airlines Co.

By Year | By Airline | By Year For Each Airline



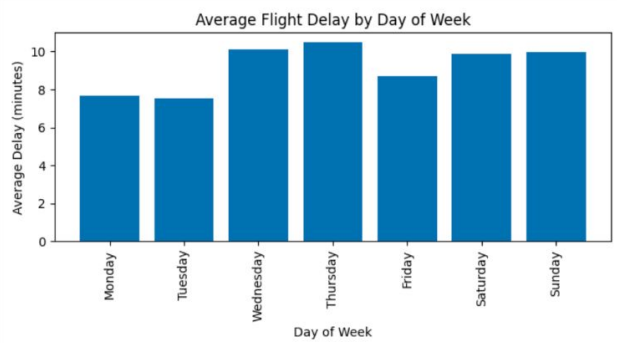
By Year | By Airline | By Year For Each Airline

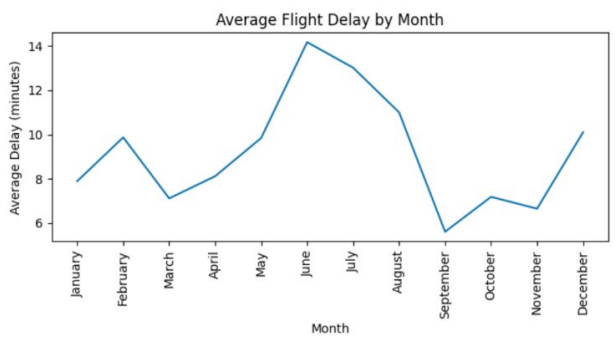


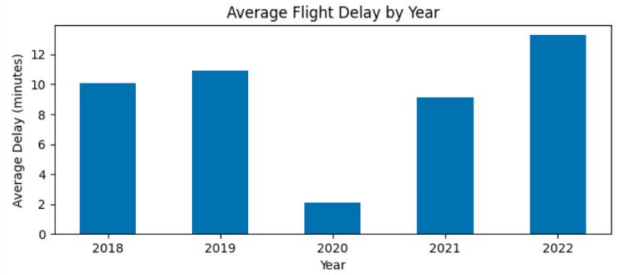

Scheduled Flights by Year for Each Airline
1e6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Flight Count
2018
2019
2020
2021
2022
Year
Air Wisconsin Airlines Corp
Alaska Airlines Inc.
Allegiant Air
American Airlines Inc.
Cape Air
Capital Cargo International
Comair Inc.
Commutair Aka Champlain Enterprises, Inc.
Compass Airlines
Delta Air Lines Inc.
Empire Airlines Inc.
Endeavor Air Inc.
Envoy Air
ExpressJet Airlines Inc.
Frontier Airlines Inc.
GoJet Airlines, LLC d/b/a United Express
Hawaiian Airlines Inc.
Horizon Air
JetBlue Airways
Mesa Airlines Inc.
Peninsula Airways Inc.
Republic Airlines
SkyWest Airlines Inc.
Southwest Airlines Co.
Spirit Air Lines
Trans States Airlines
United Air Lines Inc.
Virgin America

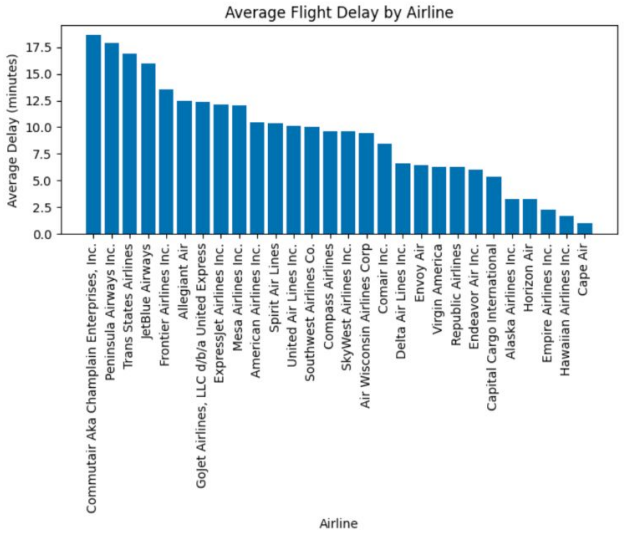
Delays | Delay Groups | Airports

Average Delay by Day of Week, Month, Year and Airline









Average Delay by Year

Show 25 entries

Search:

Year	Average Delay (in minutes)
2018	10.080462143237643
2019	10.91407491012457
2020	2.071304789630669
2021	9.154191675646333
2022	13.280008630920271

Showing 1 to 5 of 5 entries

Previous 1 Next

Average Delay by Airline

Show 25 entries

Search:

Name of the Airline	Average Delay (in minutes)
Air Wisconsin Airlines Corp	9.44
Alaska Airlines Inc.	3.29
Allegiant Air	12.45
American Airlines Inc.	10.46
Cape Air	1.0
Capital Cargo International	5.39
Comair Inc.	8.44
Commutair Aka Champlain Enterprises, Inc.	18.6
Compass Airlines	9.66
Delta Air Lines Inc.	6.62
Empire Airlines Inc.	2.25
Endeavor Air Inc.	6.04

Average Delay by Day of Week

Show 25 entries Search:

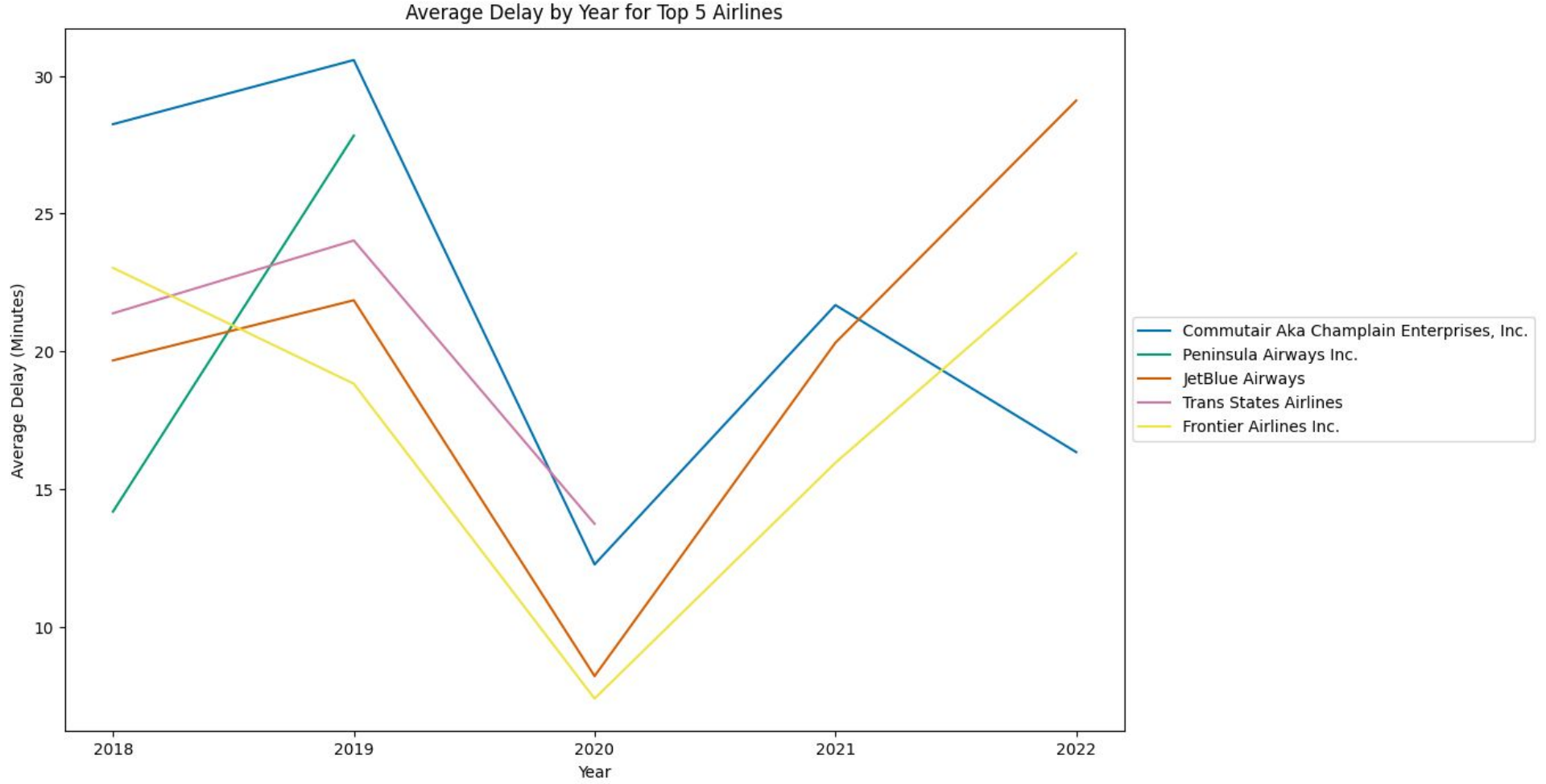
Day of Week	Average Delay (in minutes)
Friday	8.71
Monday	7.69
Saturday	9.89
Sunday	9.99
Thursday	10.47
Tuesday	7.53
Wednesday	10.13

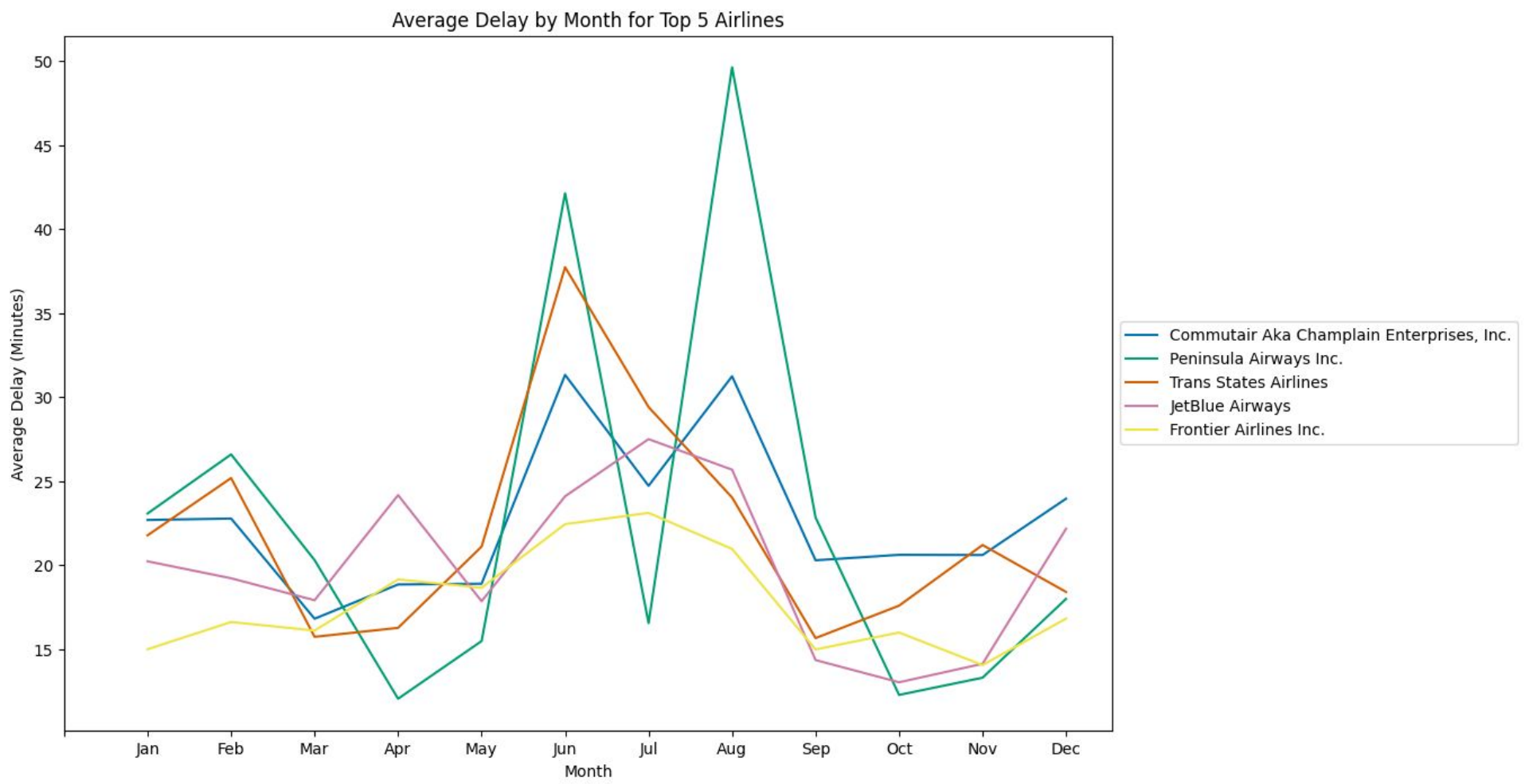
Showing 1 to 7 of 7 entries Previous 1 Next

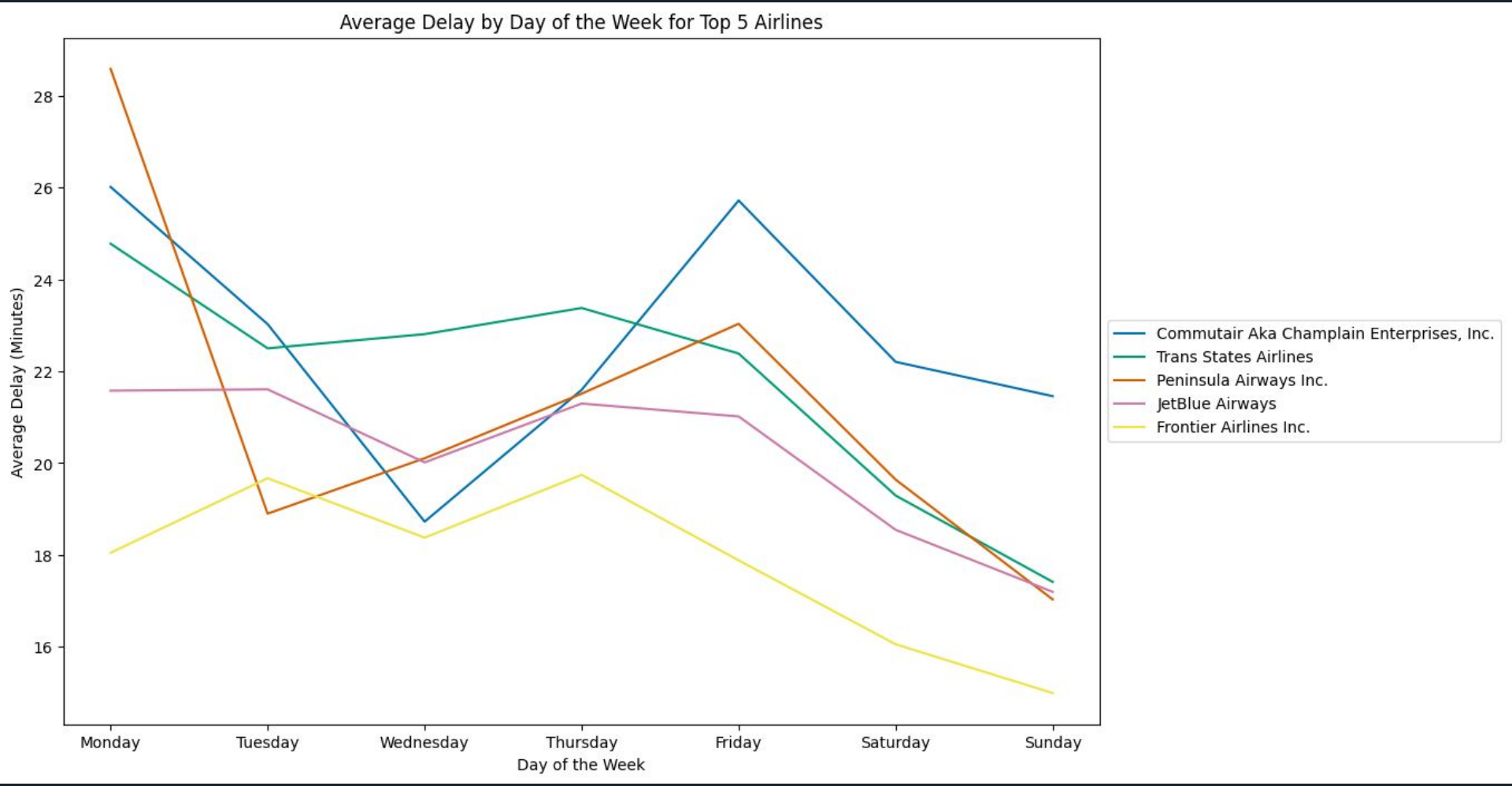
Average Delay by Month

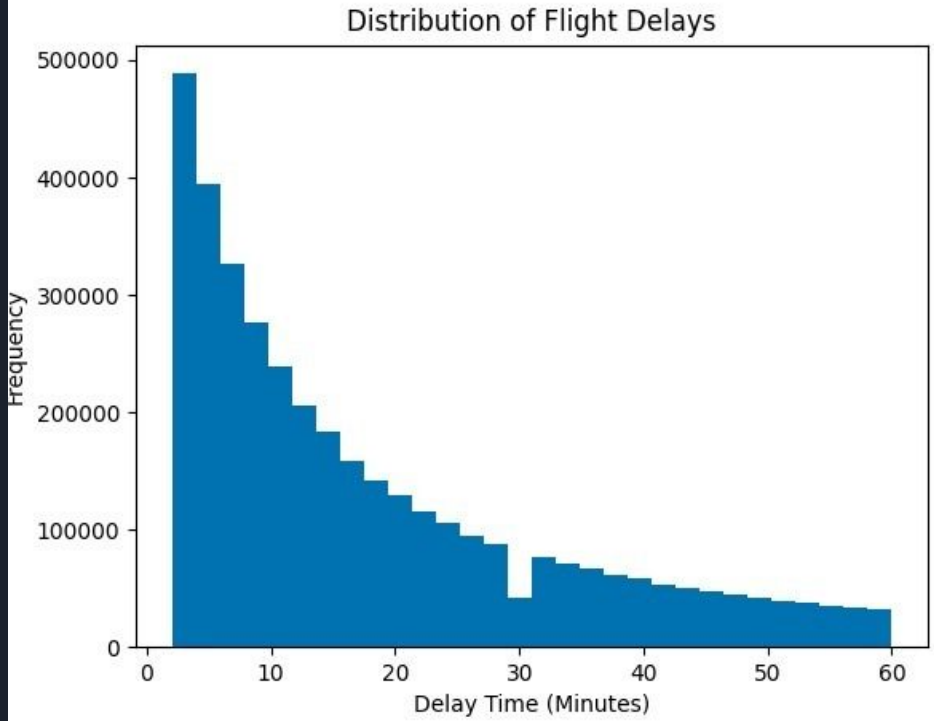
Show 25 entries Search:

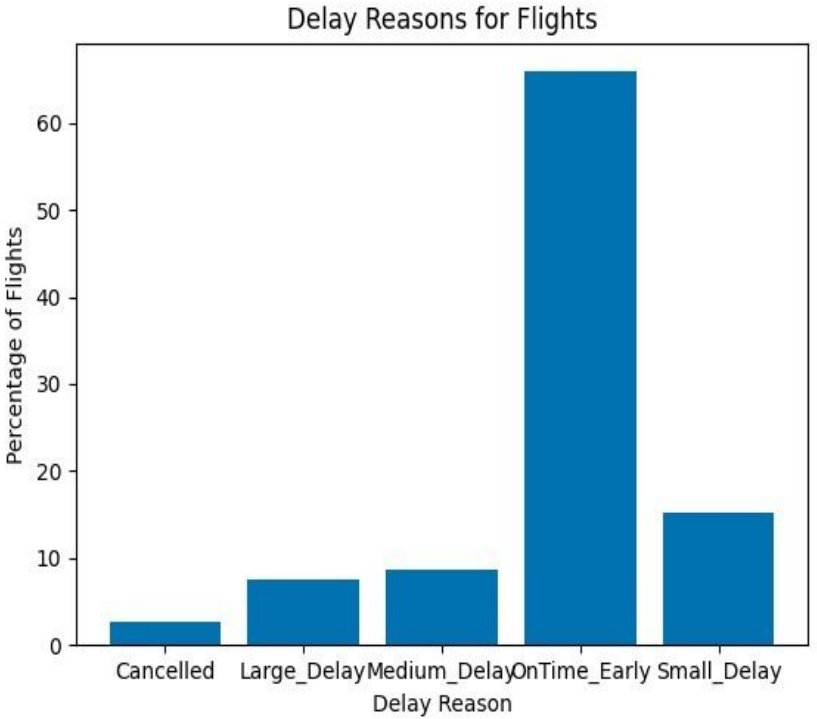
Month	Average Delay (in minutes)
April	8.11
August	11.0
December	10.1
February	9.87
January	7.89
July	13.02
June	14.18
March	7.11
May	9.84
November	6.64


Average Delay by Year for Top 5 Airlines
Average Delay (Minutes)
30
25
20
15
10
2018
2019
2020
2021
2022
Year
Commutair Aka Champlain Enterprises, Inc.
Peninsula Airways Inc.
JetBlue Airways
Trans States Airlines
Frontier Airlines Inc.


Average Delay by Month for Top 5 Airlines
Average Delay (Minutes)
50
45
40
35
30
25
20
15
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
Commutair Aka Champlain Enterprises, Inc.
Peninsula Airways Inc.
Trans States Airlines
JetBlue Airways
Frontier Airlines Inc.


Average Delay by Day of the Week for Top 5 Airlines
Average Delay (Minutes)
28
26
24
22
20
18
16
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the Week
Commutair Aka Champlain Enterprises, Inc.
Trans States Airlines
Peninsula Airways Inc.
JetBlue Airways
Frontier Airlines Inc.


Distribution of Flight Delays
Frequency
500000
400000
300000
200000
100000
0
0
10
20
30
40
50
60
Delay Time (Minutes)

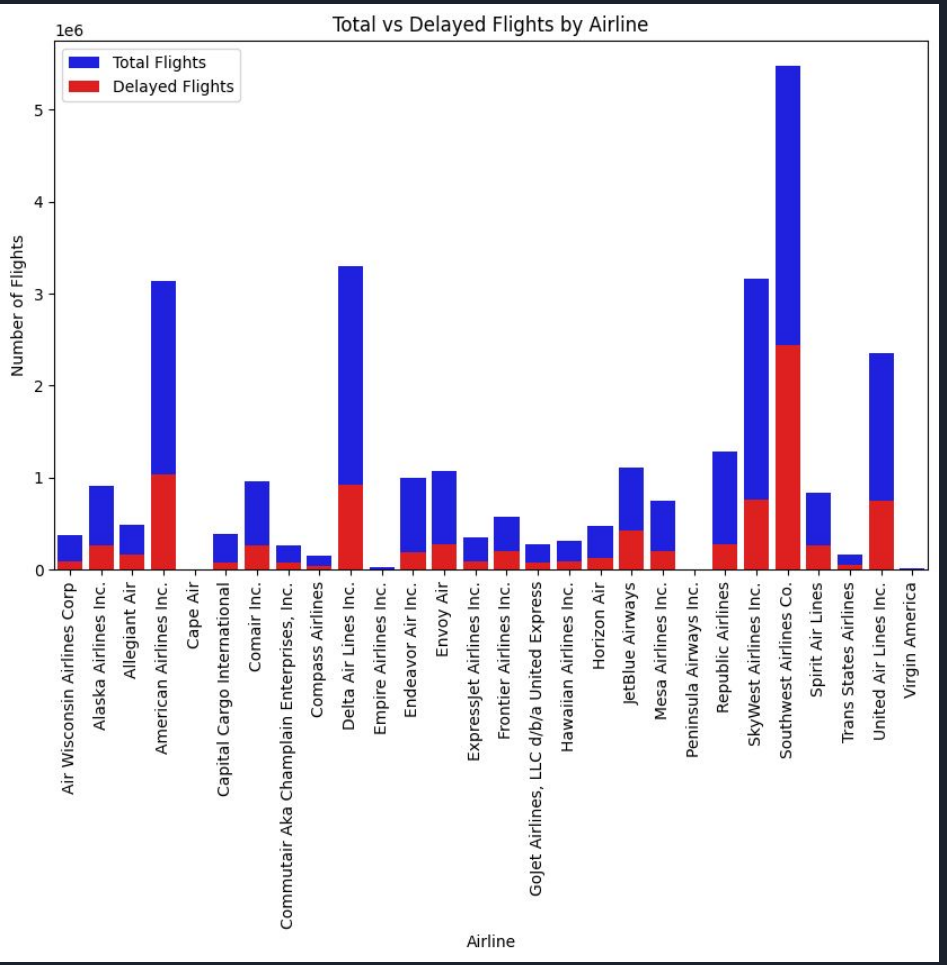

Delay Reasons for Flights
Percentage of Flights
60
50
40
30
20
10
0
Cancelled
Large_Delay
Medium_Delay
OnTime_Early
Small_Delay
Delay Reason

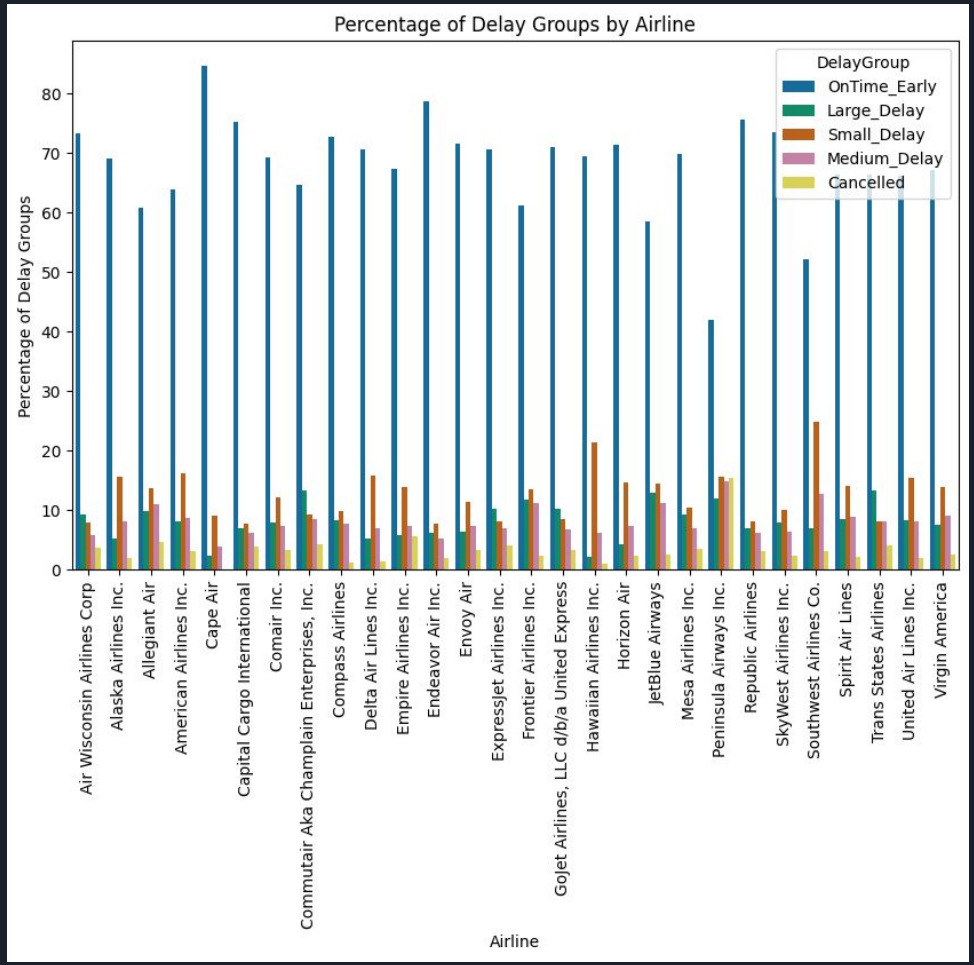
Percentage of Flight Results by Year

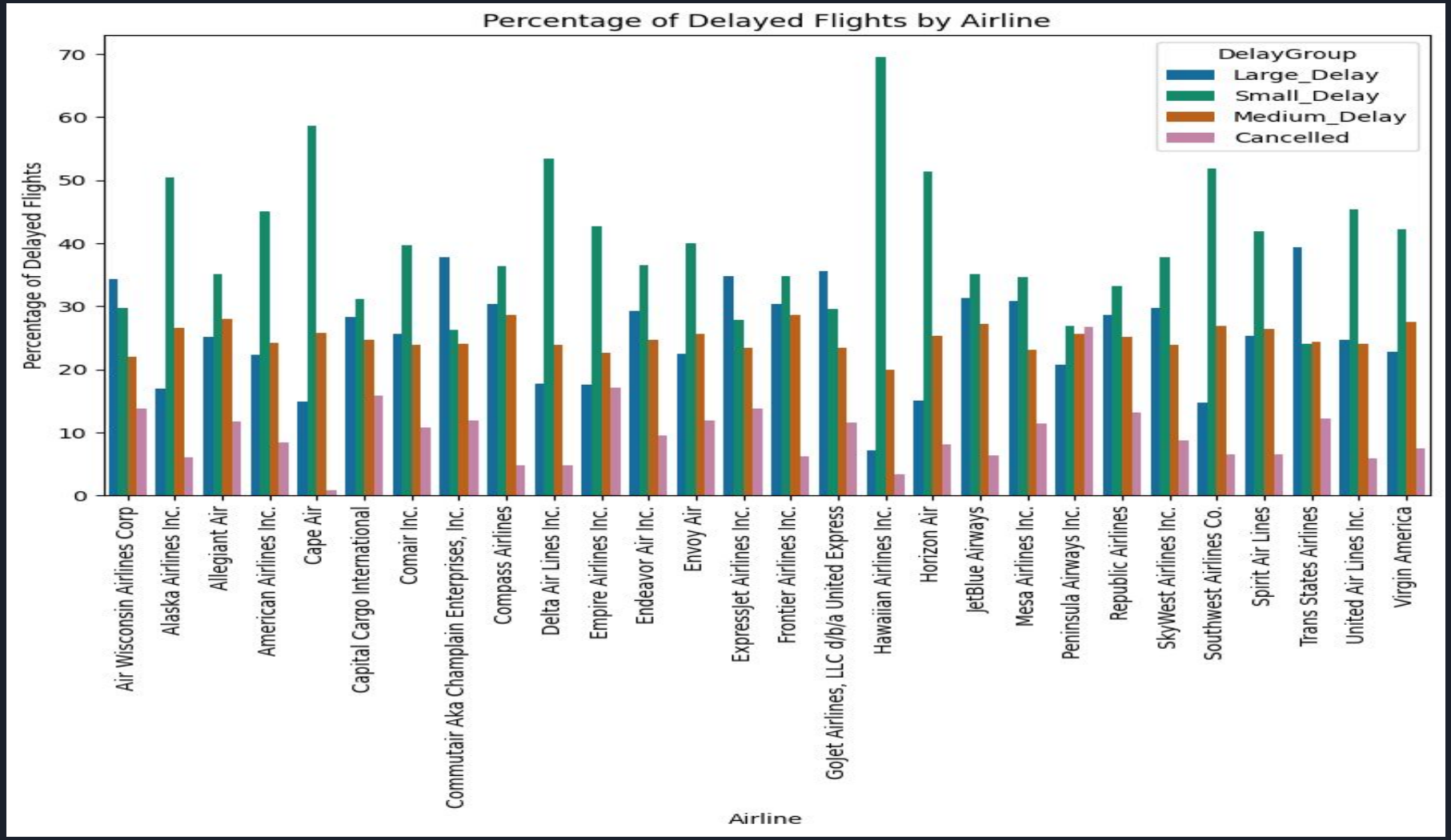
DelayGroup	OnTime_Early	Small_Delay	Medium_Delay	Large_Delay	Cancelled
Year					
2018	64.115615	16.474583	9.538448	8.317789	1.553565
2019	64.619787	15.716778	9.081583	8.683248	1.898604
2020	76.052291	9.653418	4.572319	3.727722	5.994249
2021	65.698142	16.145308	8.917879	7.479795	1.758876
2022	58.801153	17.706368	10.926539	9.545283	3.020657

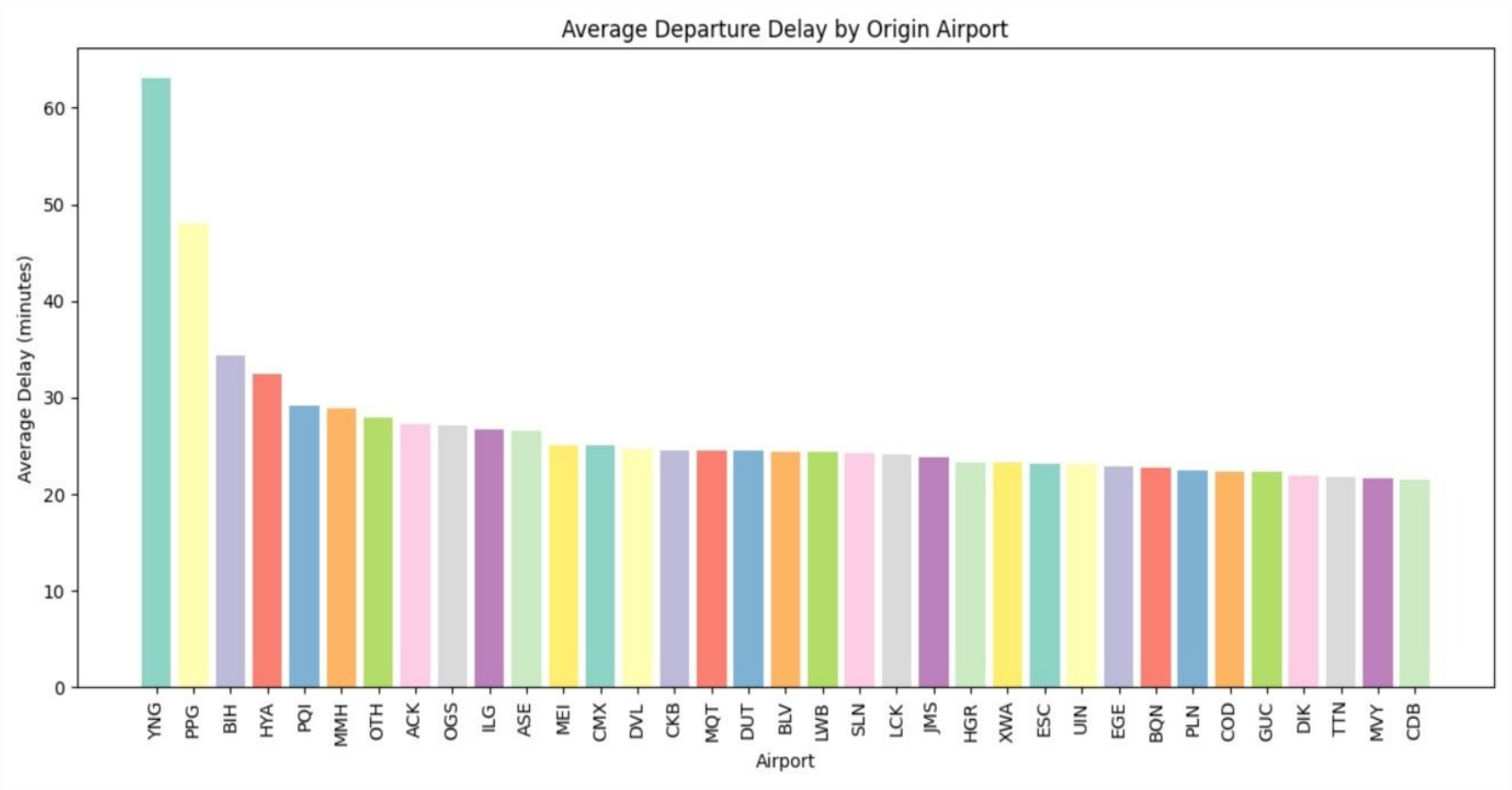
Percentage of Flight Results by Month

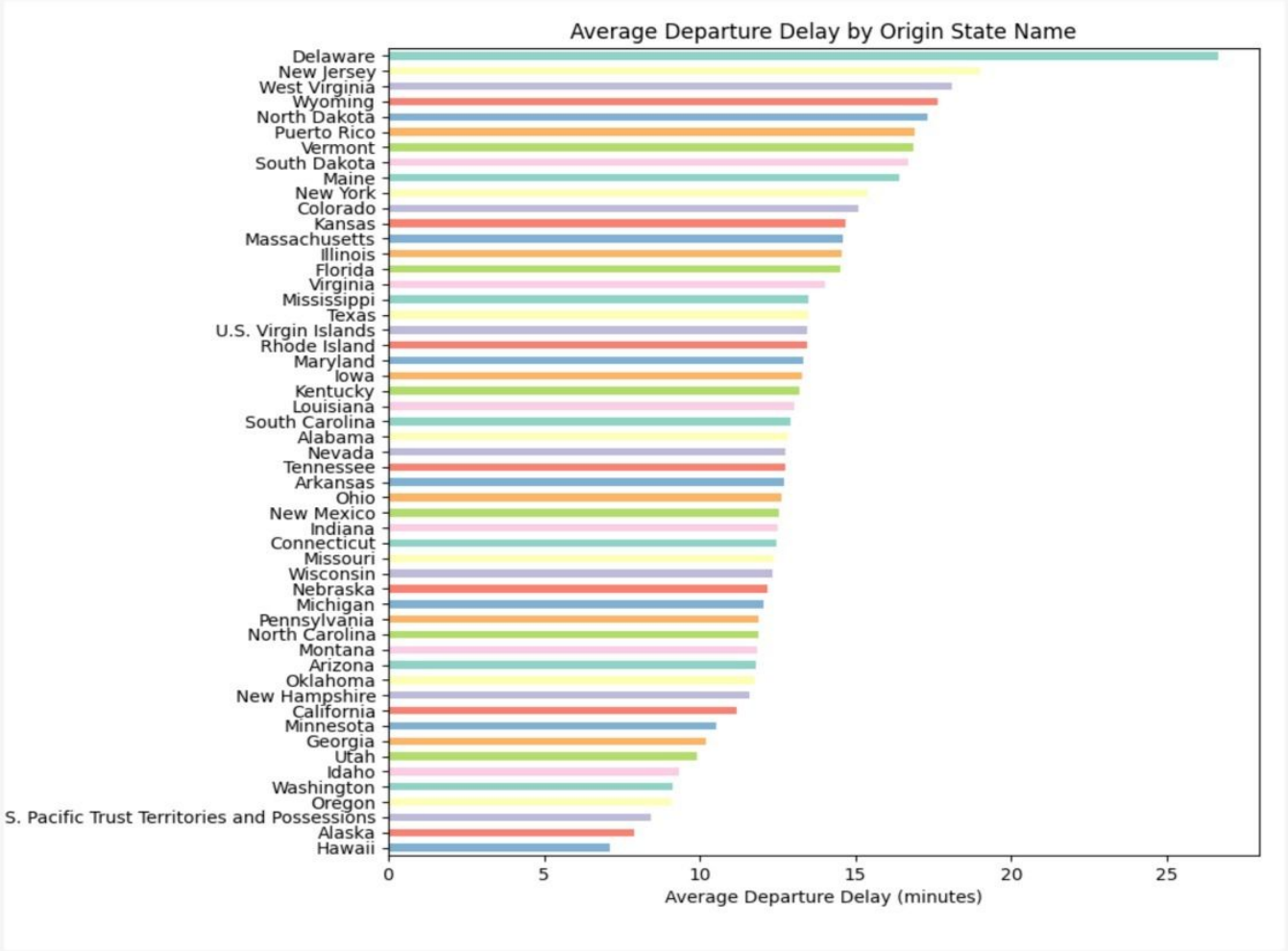
DelayGroup	OnTime_Early	Small_Delay	Medium_Delay	Large_Delay	Cancelled
Month					
April	64.436275	14.051844	7.824258	6.853866	6.833756
August	64.622096	15.009025	9.096512	9.137345	2.135022
December	64.138827	16.782343	9.731652	7.916722	1.430456
February	64.960611	15.300969	8.828322	7.786858	3.123240
January	68.169907	13.961832	7.878093	7.008976	2.981192
July	61.996702	16.361900	10.052315	9.906083	1.683000
June	59.488265	17.349791	10.754222	10.442093	1.965629
March	66.363322	14.172305	7.716007	6.173113	5.575252
May	65.023841	16.224252	8.965554	7.875154	1.911198
November	70.182986	15.208750	7.826532	5.941371	0.840361
October	69.570264	14.859330	7.970955	6.487423	1.112028
September	72.924881	13.140776	6.725733	5.813787	1.394822


Total vs Delayed Flights by Airline
1e6
Total Flights
Delayed Flights
Number of Flights
0
1
2
3
4
5
Air Wisconsin Airlines Corp
Alaska Airlines Inc.
Allegiant Air
American Airlines Inc.
Cape Air
Capital Cargo International
Comair Inc.
Commutair Aka Champlain Enterprises, Inc.
Compass Airlines
Delta Air Lines Inc.
Empire Airlines Inc.
Endeavor Air Inc.
Envoy Air
ExpressJet Airlines Inc.
Frontier Airlines Inc.
Gojet Airlines, LLC d/b/a United Express
Hawaiian Airlines Inc.
Horizon Air
JetBlue Airways
Mesa Airlines Inc.
Peninsula Airways Inc.
Republic Airlines
SkyWest Airlines Inc.
Southwest Airlines Co.
Spirit Air Lines
Trans States Airlines
United Air Lines Inc.
Virgin America
Airline


Percentage of Delay Groups by Airline
DelayGroup
OnTime_Early
Large_Delay
Small_Delay
Medium_Delay
Cancelled
Percentage of Delay Groups
0
10
20
30
40
50
60
70
80
Air Wisconsin Airlines Corp
Alaska Airlines Inc.
Allegiant Air
American Airlines Inc.
Cape Air
Capital Cargo International
Comair Inc.
Commutair Aka Champlain Enterprises, Inc.
Compass Airlines
Delta Air Lines Inc.
Empire Airlines Inc.
Endeavor Air Inc.
Envoy Air
ExpressJet Airlines Inc.
Frontier Airlines Inc.
Gojet Airlines, LLC d/b/a United Express
Hawaiian Airlines Inc.
Horizon Air
JetBlue Airways
Mesa Airlines Inc.
Peninsula Airways Inc.
Republic Airlines
SkyWest Airlines Inc.
Southwest Airlines Co.
Spirit Air Lines
Trans States Airlines
United Air Lines Inc.
Virgin America
Airline

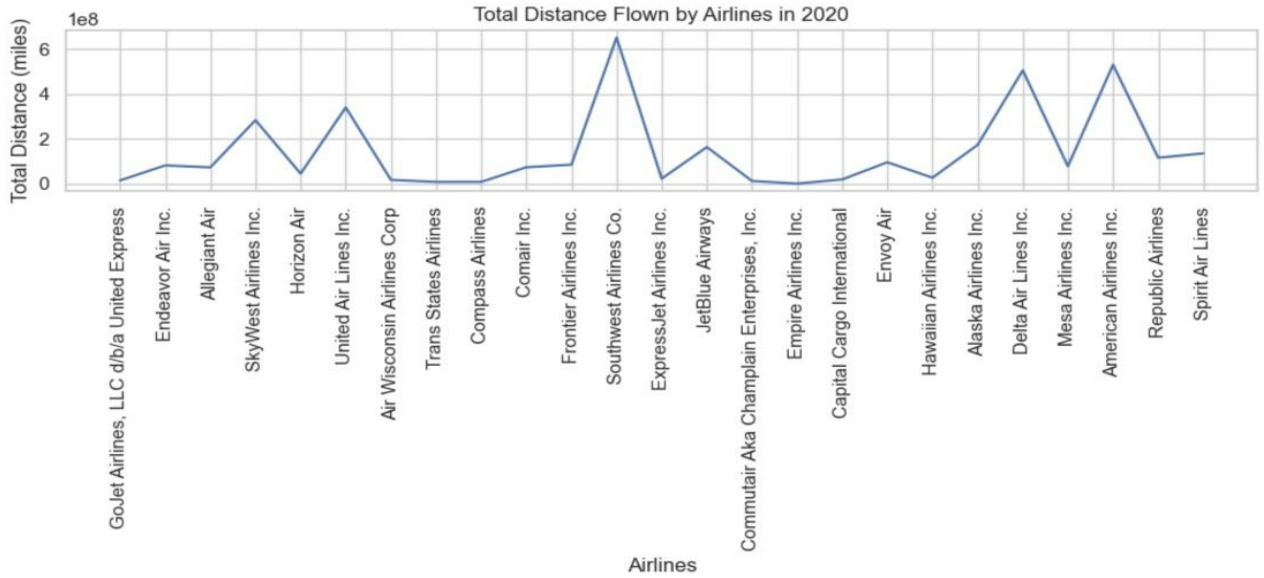

Percentage of Delayed Flights by Airline
DelayGroup
Large_Delay
Small_Delay
Medium_Delay
Cancelled
Percentage of Delayed Flights
0
10
20
30
40
50
60
70
Air Wisconsin Airlines Corp
Alaska Airlines Inc.
Allegiant Air
American Airlines Inc.
Cape Air
Capital Cargo International
Comair Inc.
Commutair Aka Champlain Enterprises, Inc.
Compass Airlines
Delta Air Lines Inc.
Empire Airlines Inc.
Endeavor Air Inc.
Envoy Air
Expressjet Airlines Inc.
Frontier Airlines Inc.
Gojet Airlines, LLC d/b/a United Express
Hawaiian Airlines Inc.
Horizon Air
JetBlue Airways
Mesa Airlines Inc.
Peninsula Airways Inc.
Republic Airlines
SkyWest Airlines Inc.
Southwest Airlines Co.
Spirit Air Lines
Trans States Airlines
United Air Lines Inc.
Virgin America
Airline


Average Departure Delay by Origin Airport
Average Delay (minutes)
0
10
20
30
40
50
60
YNG
PPG
BIH
HYA
PQI
MMH
OTH
ACK
OGS
ILG
ASE
MEI
CMX
DVL
CKB
MQT
DUT
BLV
LWB
SLN
LCK
JMS
HGR
XWA
ESC
UIN
EGE
BQN
PLN
COD
GUC
DIK
TTN
MVY
CDB
Airport



2020

Calculate



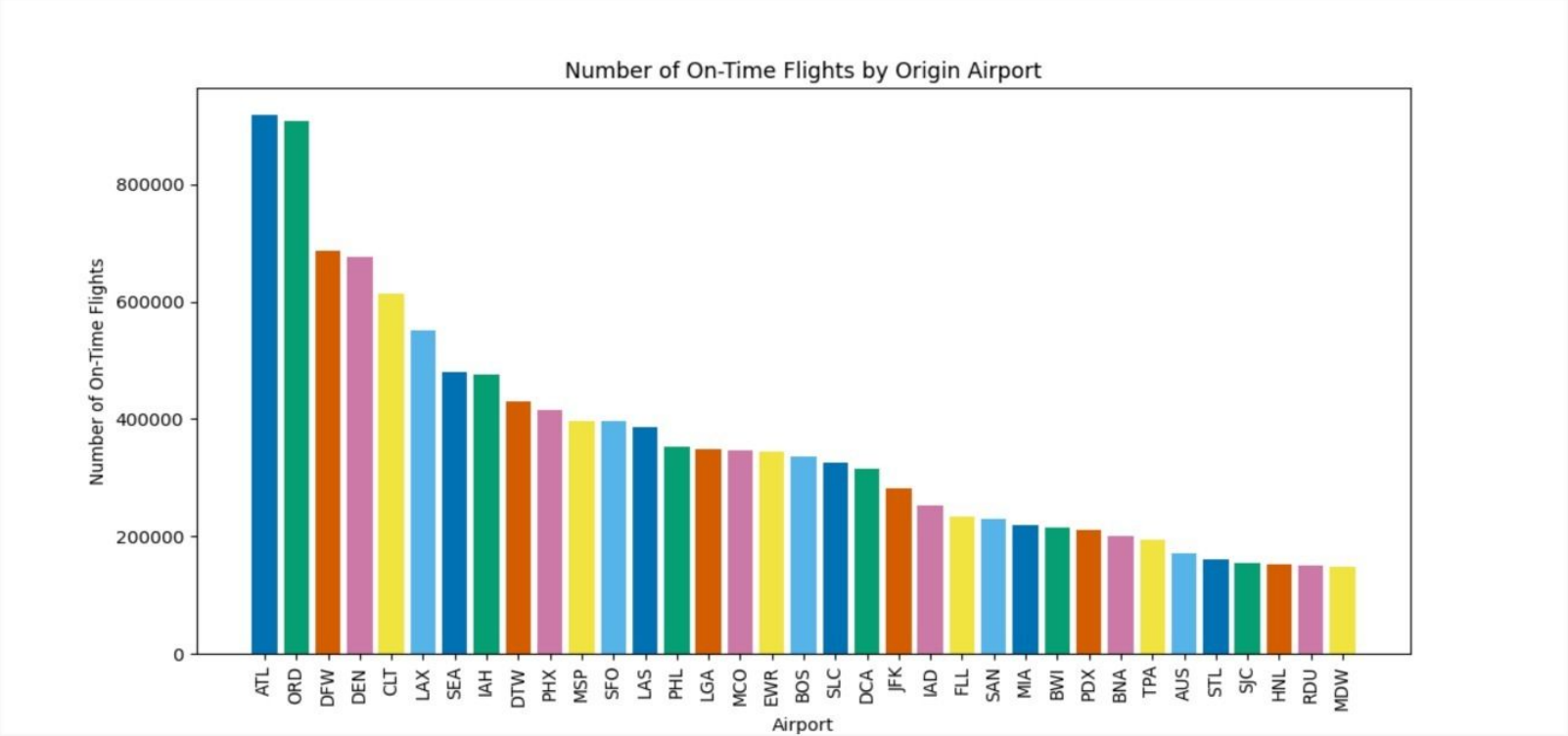
Show 25 entries

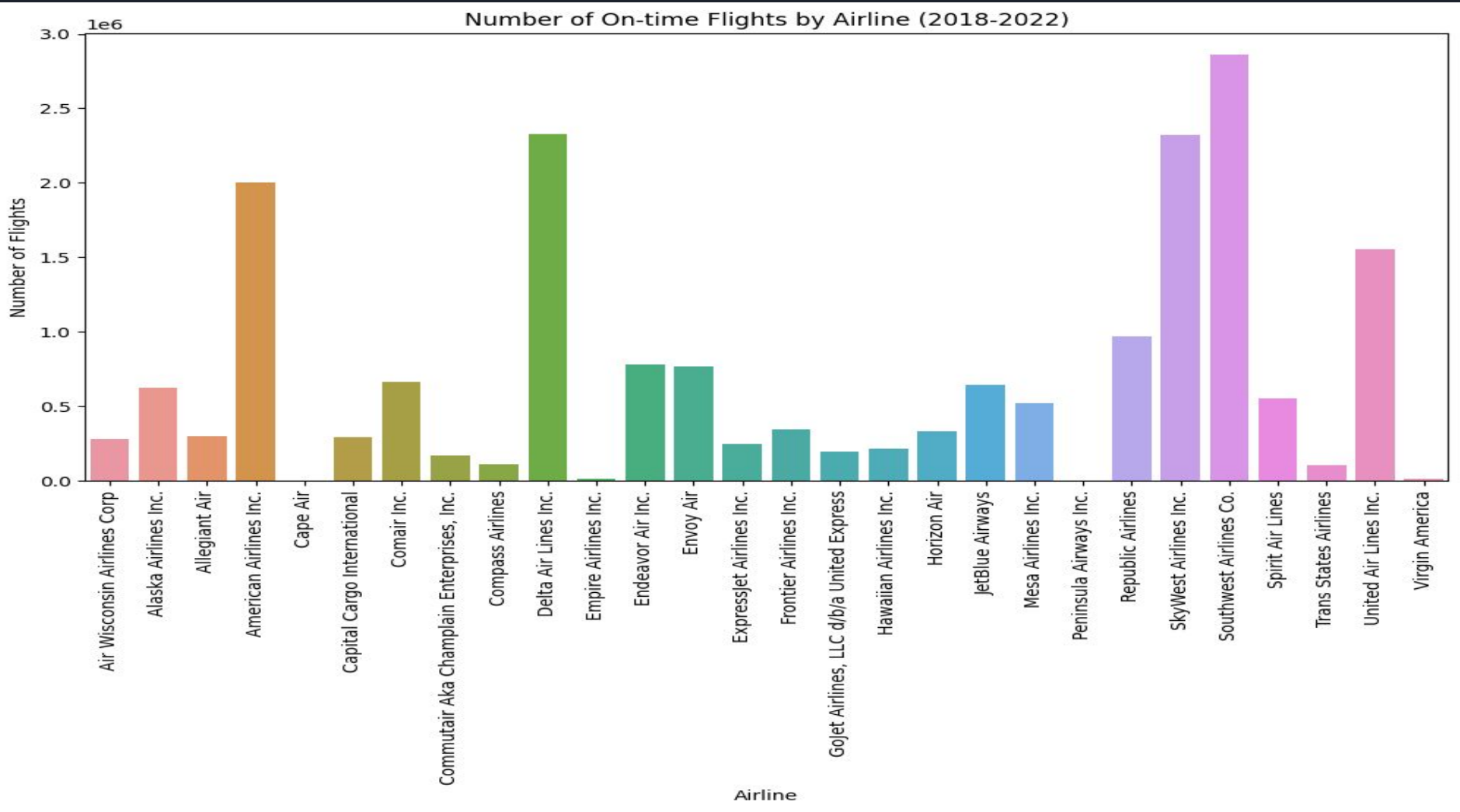
Search:

Airline Name	Distance (miles)
Air Wisconsin Airlines Corp	17098945.0
Alaska Airlines Inc.	172880381.0
Allegiant Air	72785601.0
American Airlines Inc.	529957184.0
Capital Cargo International	18948995.0
Comair Inc.	73274197.0
Commutair Aka Champlain Enterprises, Inc.	12095305.0
Compass Airlines	7977135.0
Delta Air Lines Inc.	504831955.0
Empire Airlines Inc.	310542.0
Endeavor Air Inc.	81512817.0
Envoy Air	95393191.0
ExpressJet Airlines Inc.	22287621.0
Frontier Airlines Inc.	84822494.0
GoJet Airlines, LLC d/b/a United Express	14619615.0
Hawaiian Airlines Inc.	26602425.0
Horizon Air	44850674.0
JetBlue Airways	163146080.0
Mesa Airlines Inc.	77474015.0
Republic Airlines	115623006.0

On-Time Cancelled

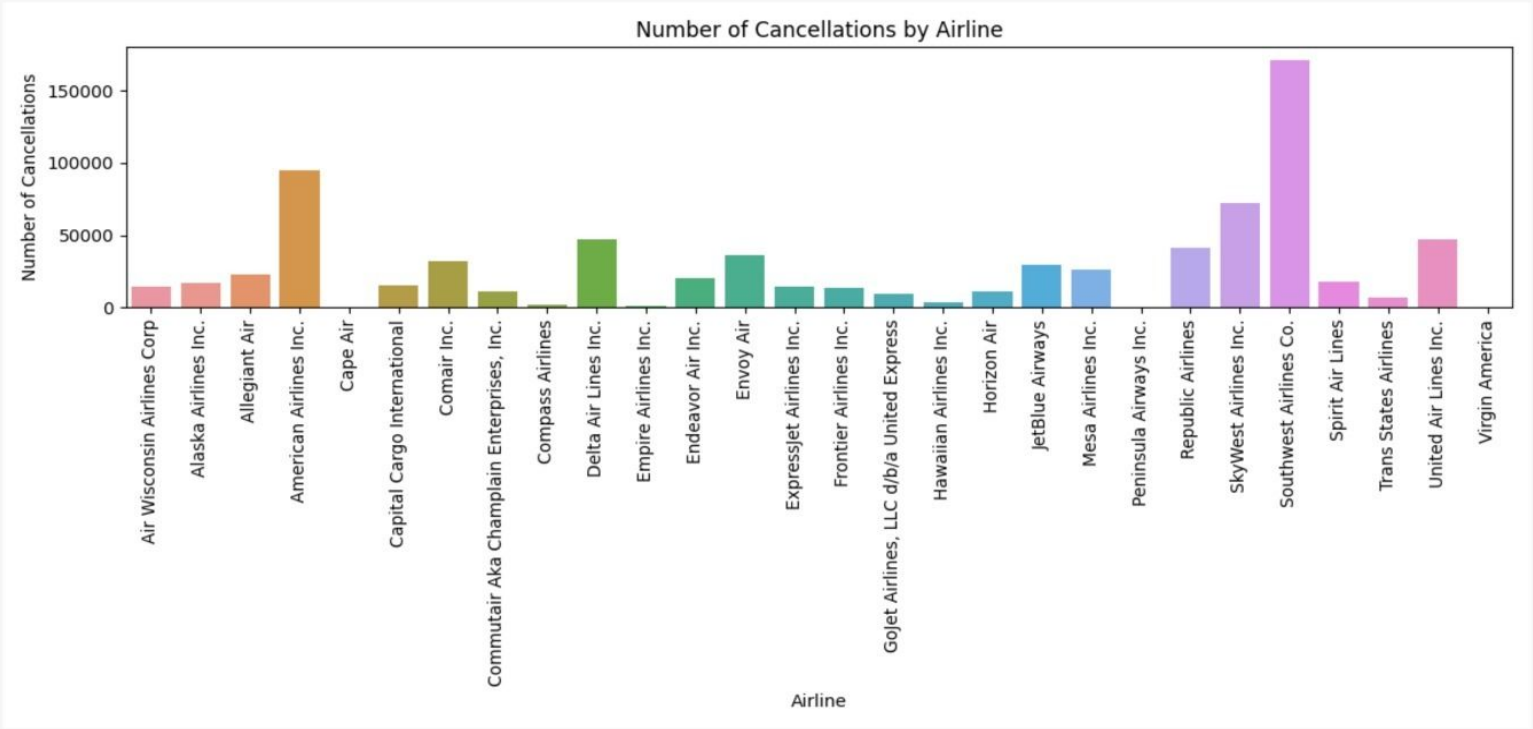
Number of On-Time Flights

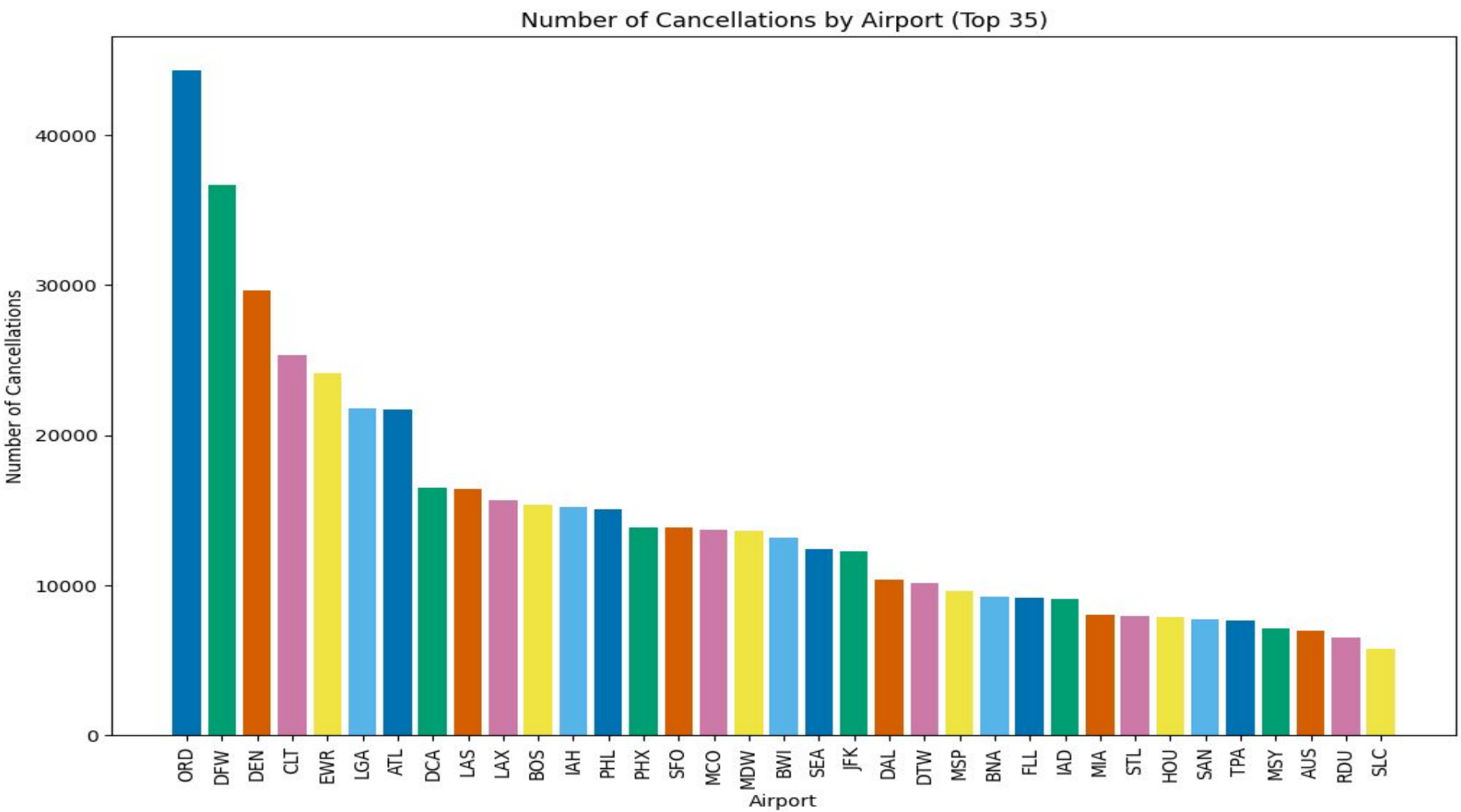



Number of On-time Flights by Airline (2018-2022)
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Number of Flights
Air Wisconsin Airlines Corp
Alaska Airlines Inc.
Allegiant Air
American Airlines Inc.
Cape Air
Capital Cargo International
Comair Inc.
Commutair Aka Champlain Enterprises, Inc.
Compass Airlines
Delta Air Lines Inc.
Empire Airlines Inc.
Endeavor Air Inc.
Envoy Air
ExpressJet Airlines Inc.
Frontier Airlines Inc.
Golet Airlines, LLC d/b/a United Express
Hawaiian Airlines Inc.
Horizon Air
JetBlue Airways
Mesa Airlines Inc.
Peninsula Airways Inc.
Republic Airlines
SkyWest Airlines Inc.
Southwest Airlines Co.
Spirit Air Lines
Trans States Airlines
United Air Lines Inc.
Virgin America
Airline

On-Time Cancelled

Number of Cancellations by Airline




Number of Cancellations by Airport (Top 35)
Number of Cancellations
0
10000
20000
30000
40000
ORD
DFW
DEN
CLT
EWR
LGA
ATL
DCA
LAS
LAX
BOS
IAH
PHL
PHX
SFO
MCO
MDW
BWI
SEA
JFK
DAL
DTW
MSP
BNA
FLL
IAD
MIA
STL
HOU
SAN
TPA
MSY
AUS
RDU
SLC
Airport

Conclusion & Future Scope

- Performed analysis on the average delays of all the flights and the flights in different delay groups.
- Analyzed the airports to see which locations had more on-time flights and cancellations.
- In order to enhance the accuracy and quick access of our results, we can use a real-time dataset for our analysis.
- Develop predictive models for delays and cancellations.