

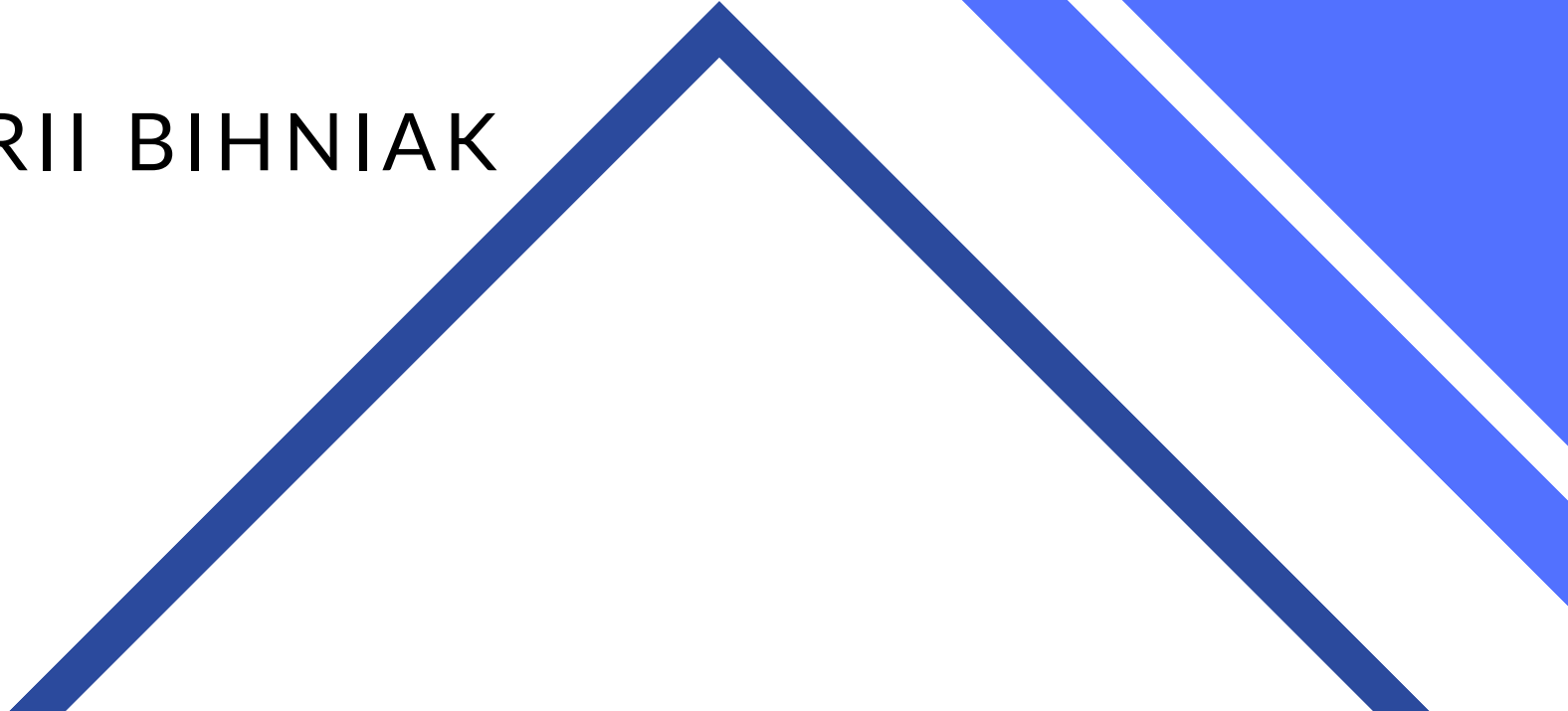


**STRUCTURED DATA **  
**PROCESSING**

# **AIRLINE PERFORMANCE**



PAWEŁ POZORSKI, IGOR KOŁODZIEJ, NAZARII BIHNIAK



3

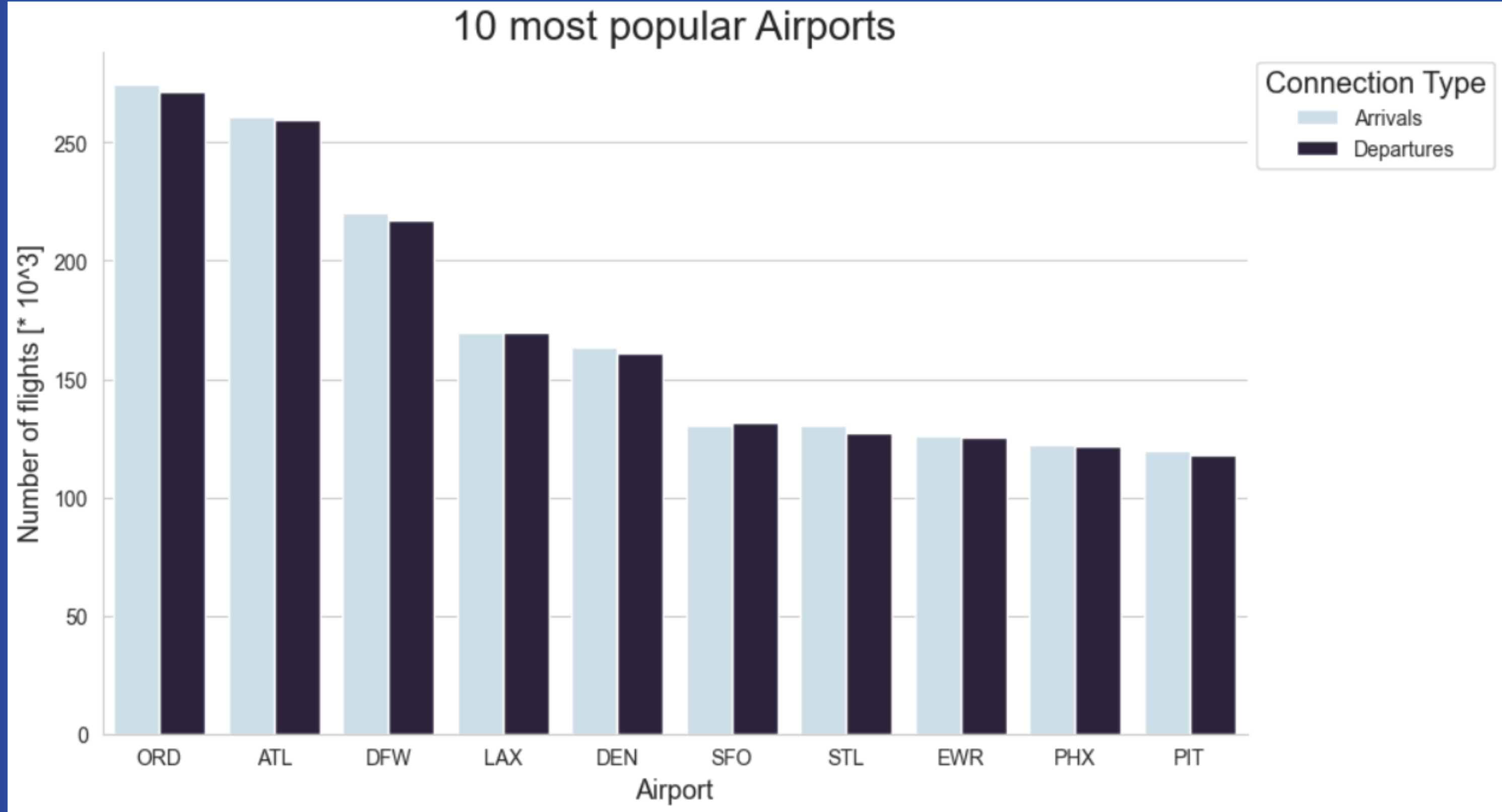
# PROJECT GOAL

*AMERICAN STATISTICAL ASSOCIATION DATA  
EXPO 2009: AIRLINE ON-TIME PERFORMANCE*

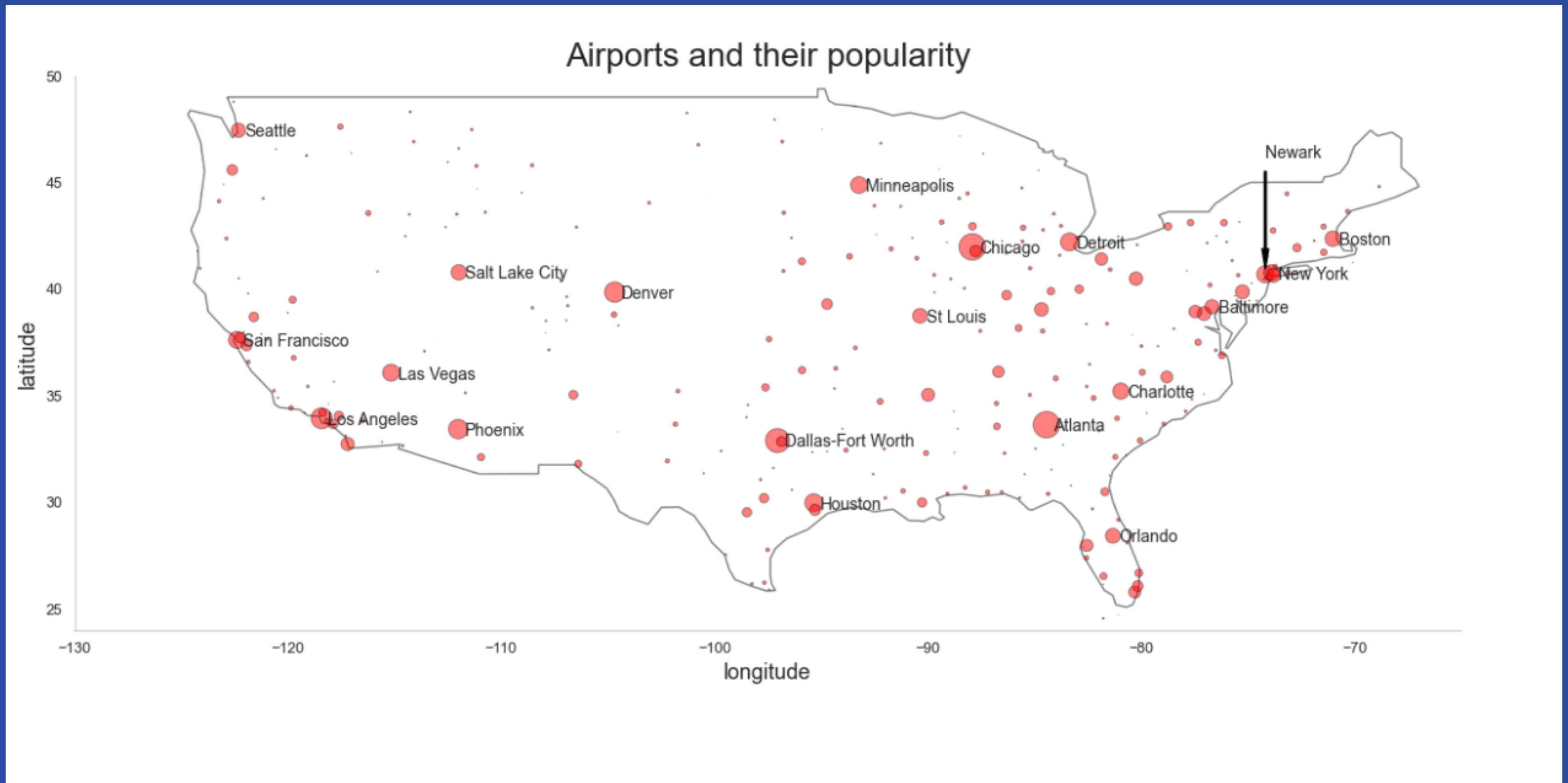


**LET'S START  
WITH CHARTS!**

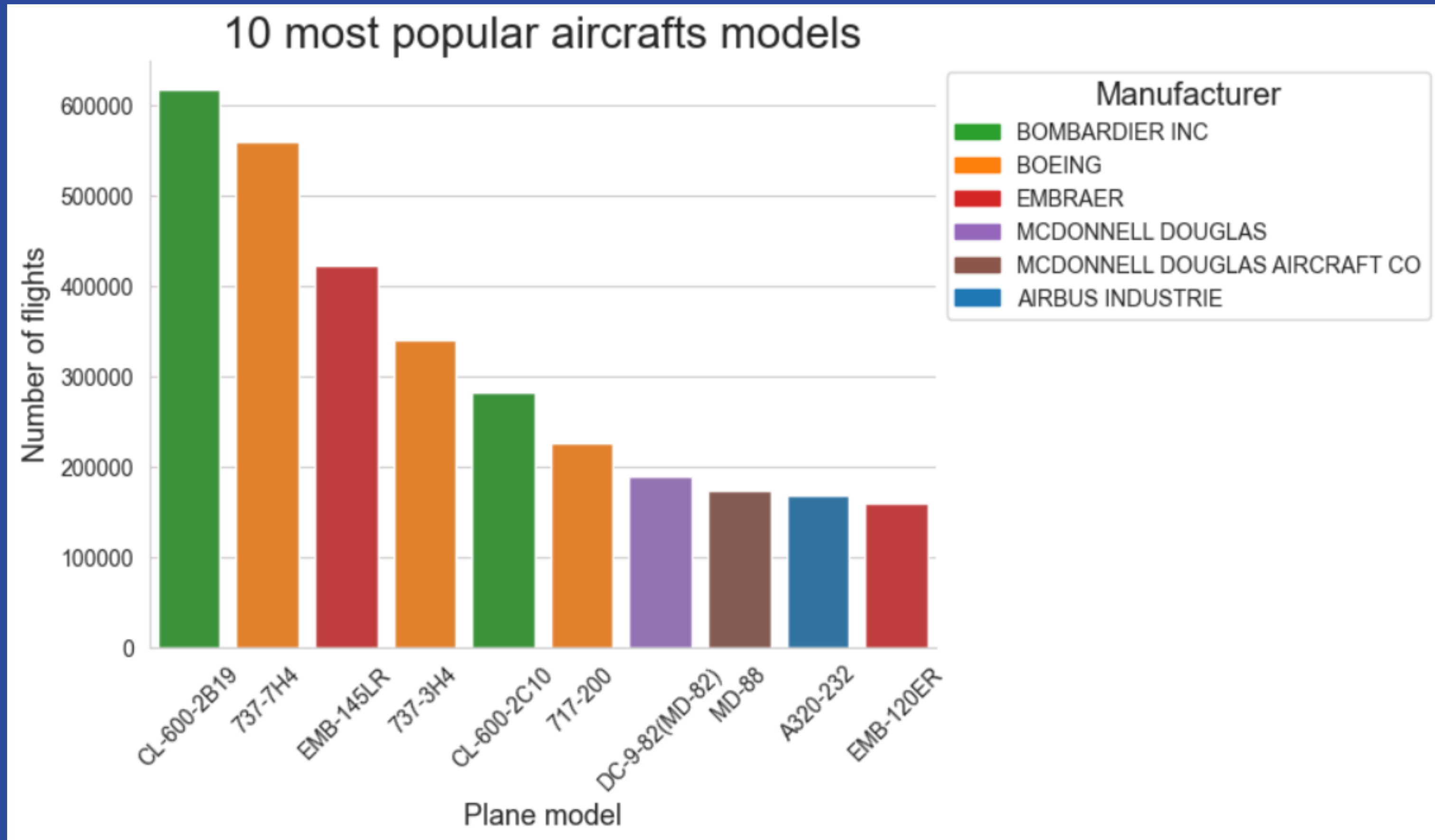
# POPULARITY



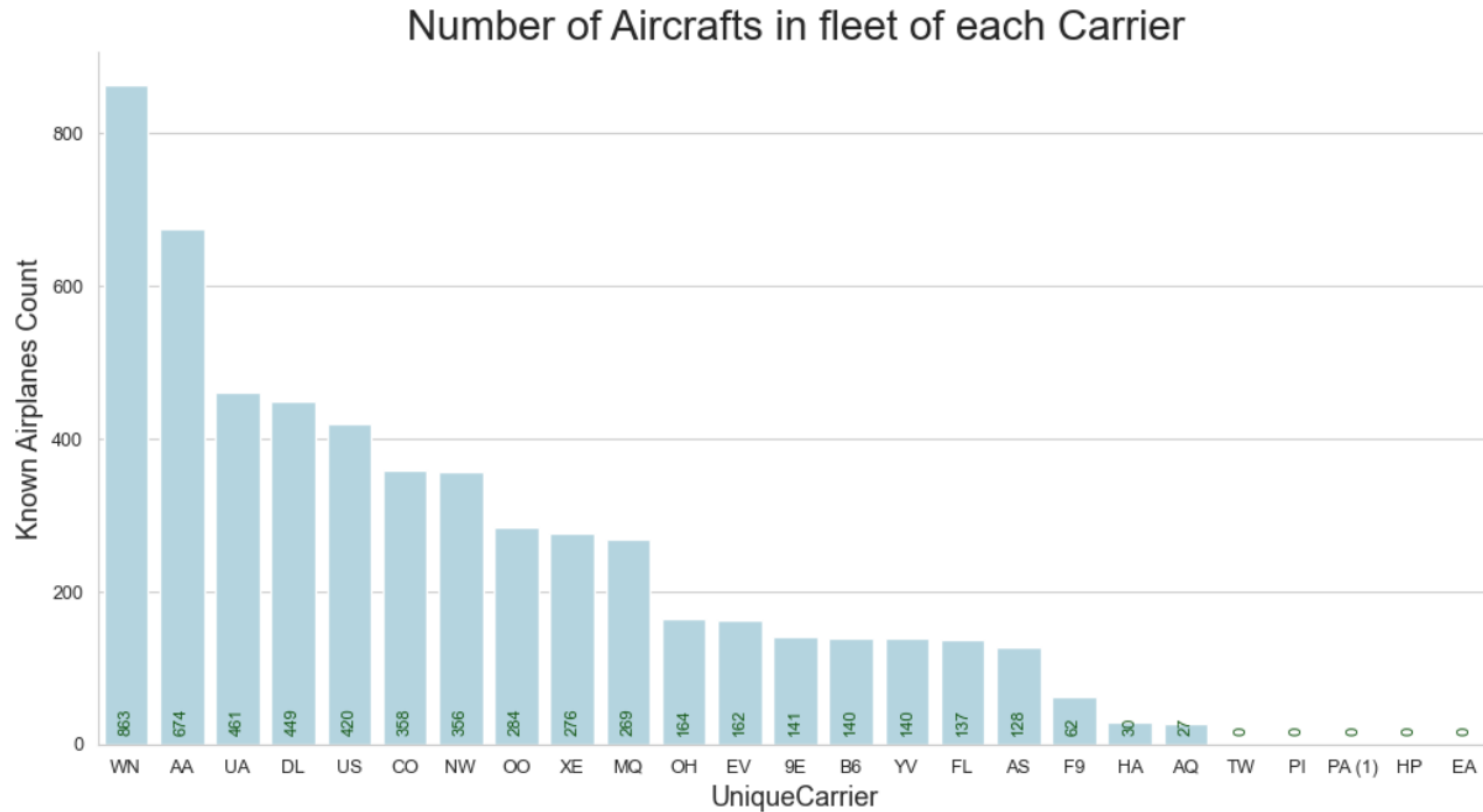
# POPULARITY



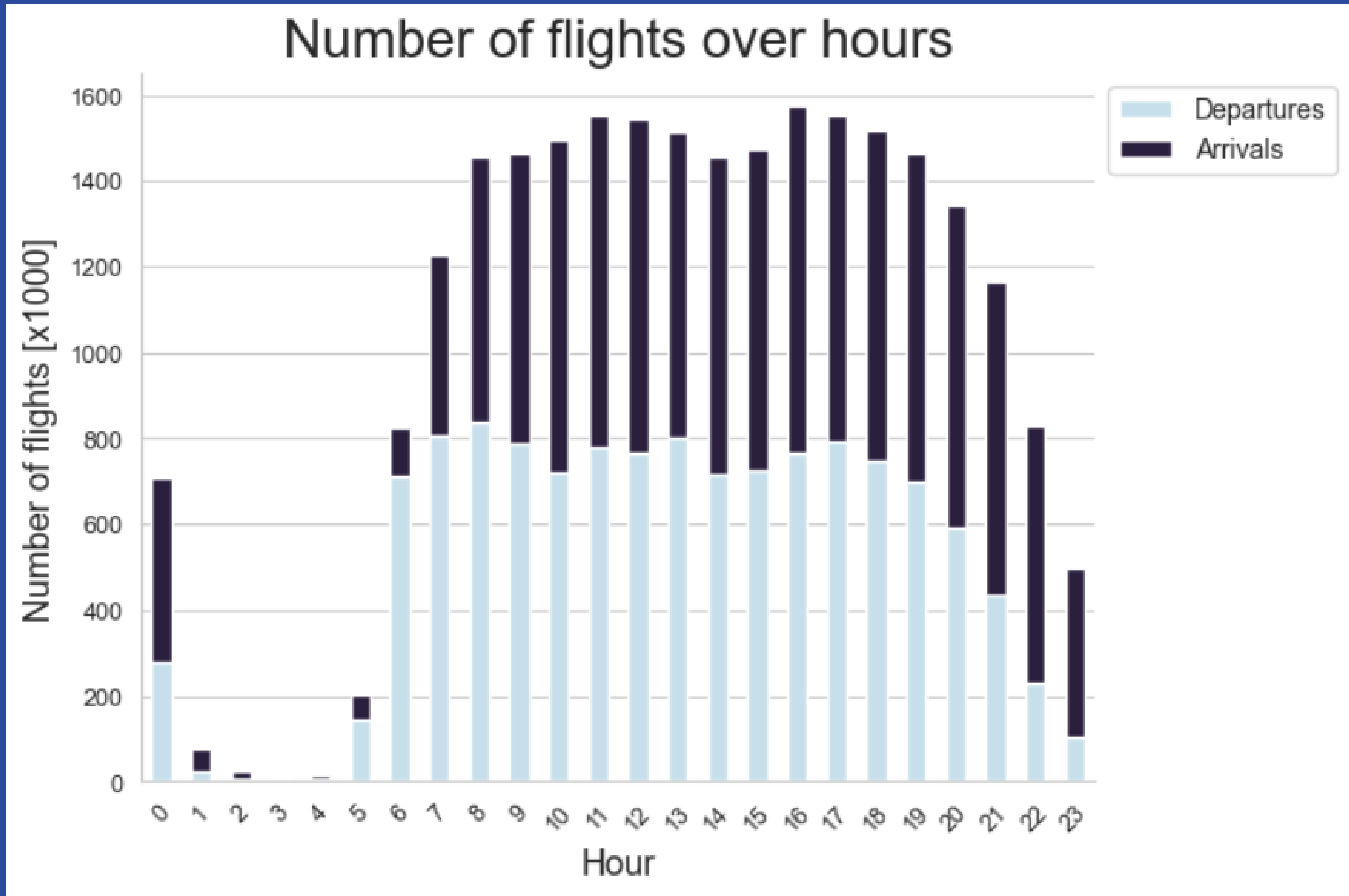
# POPULARITY



# POPULARITY



# POPULARITY



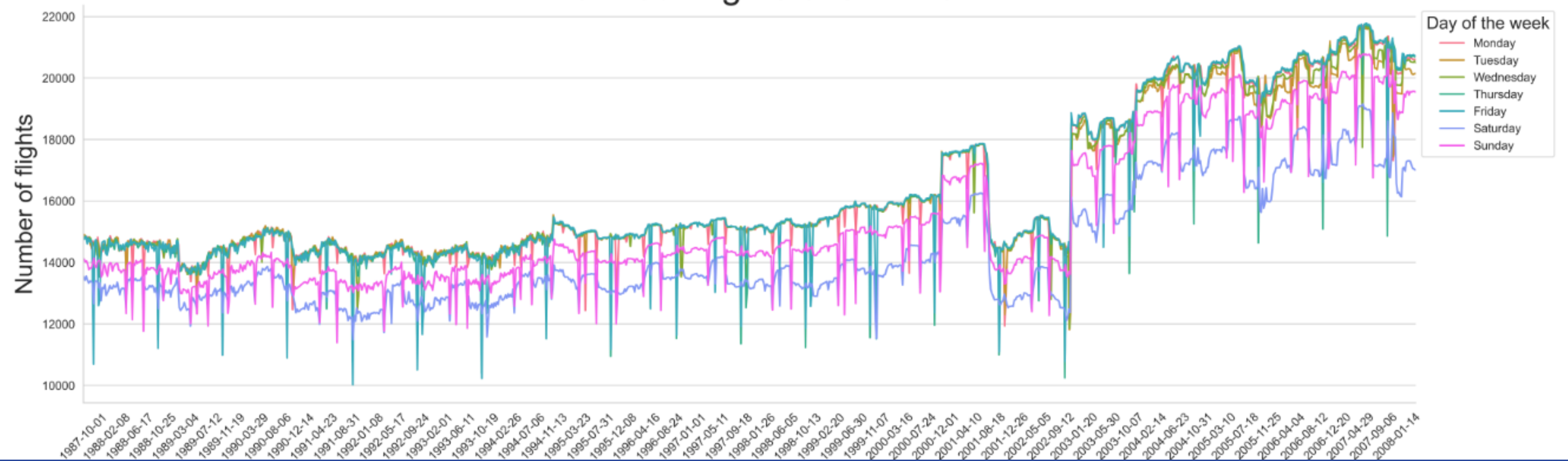


# POPULARITY

## Most popular routes

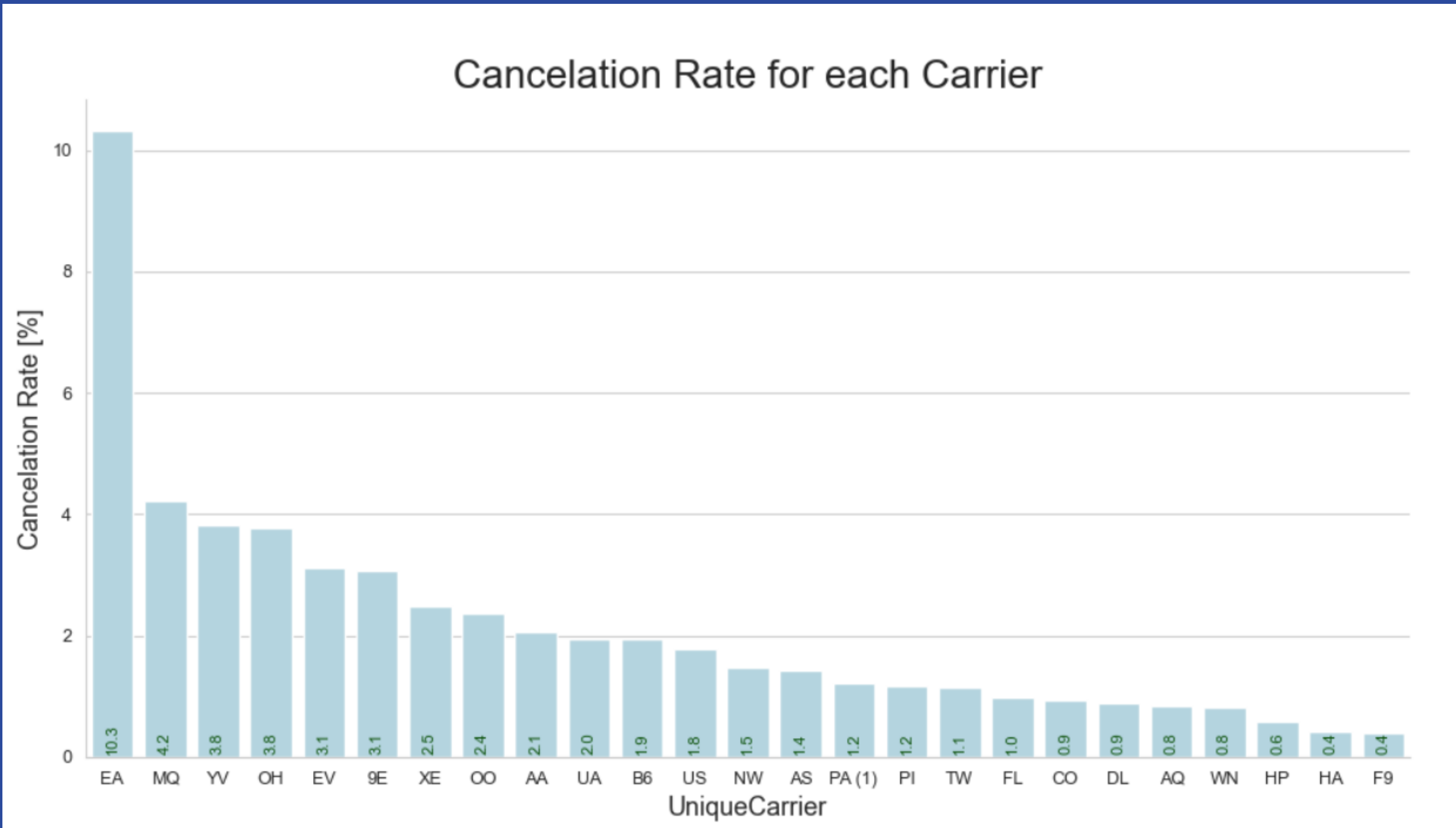


## Planned Flights over Time

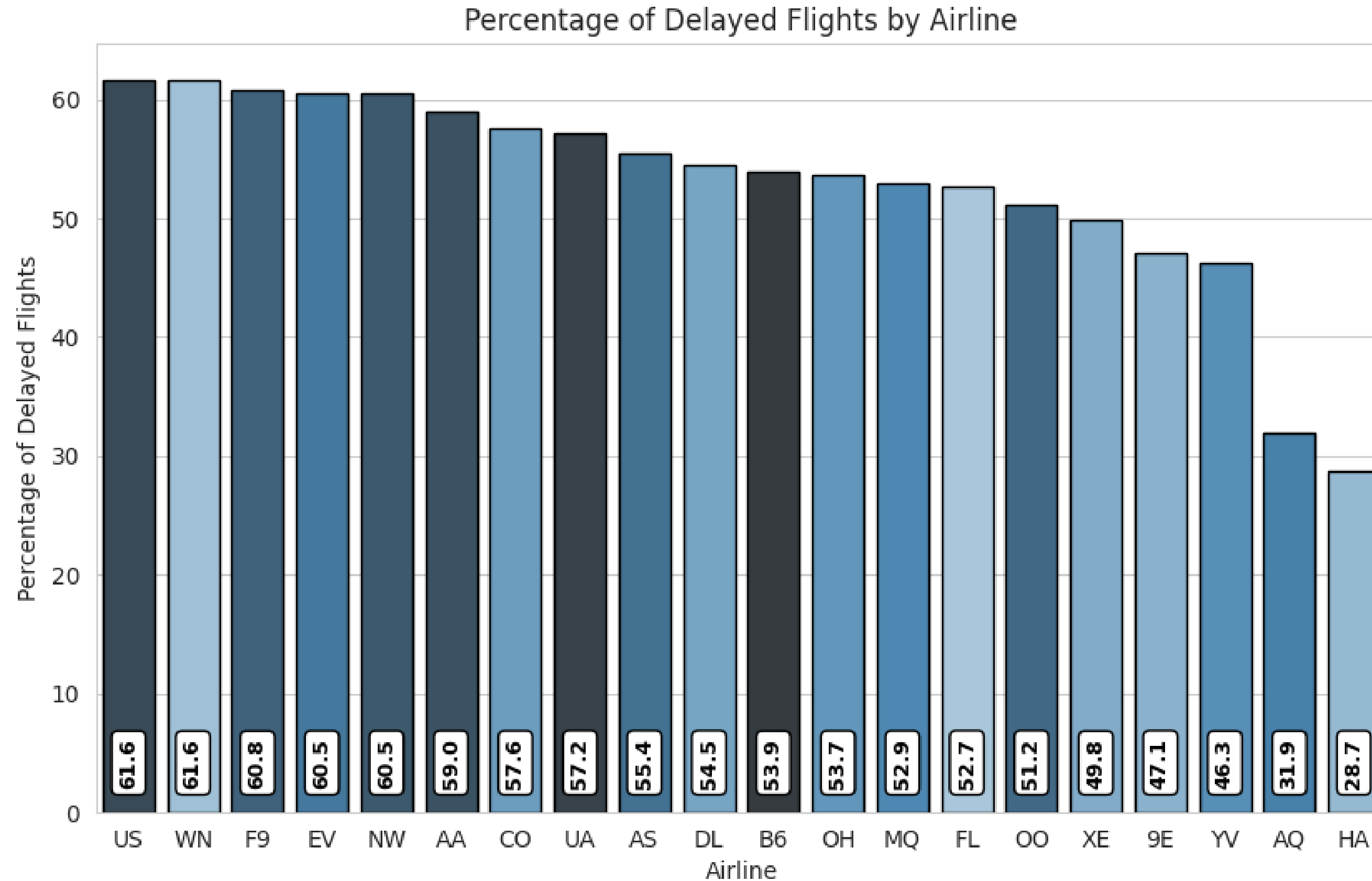


**IS IT ALL GOOD?**

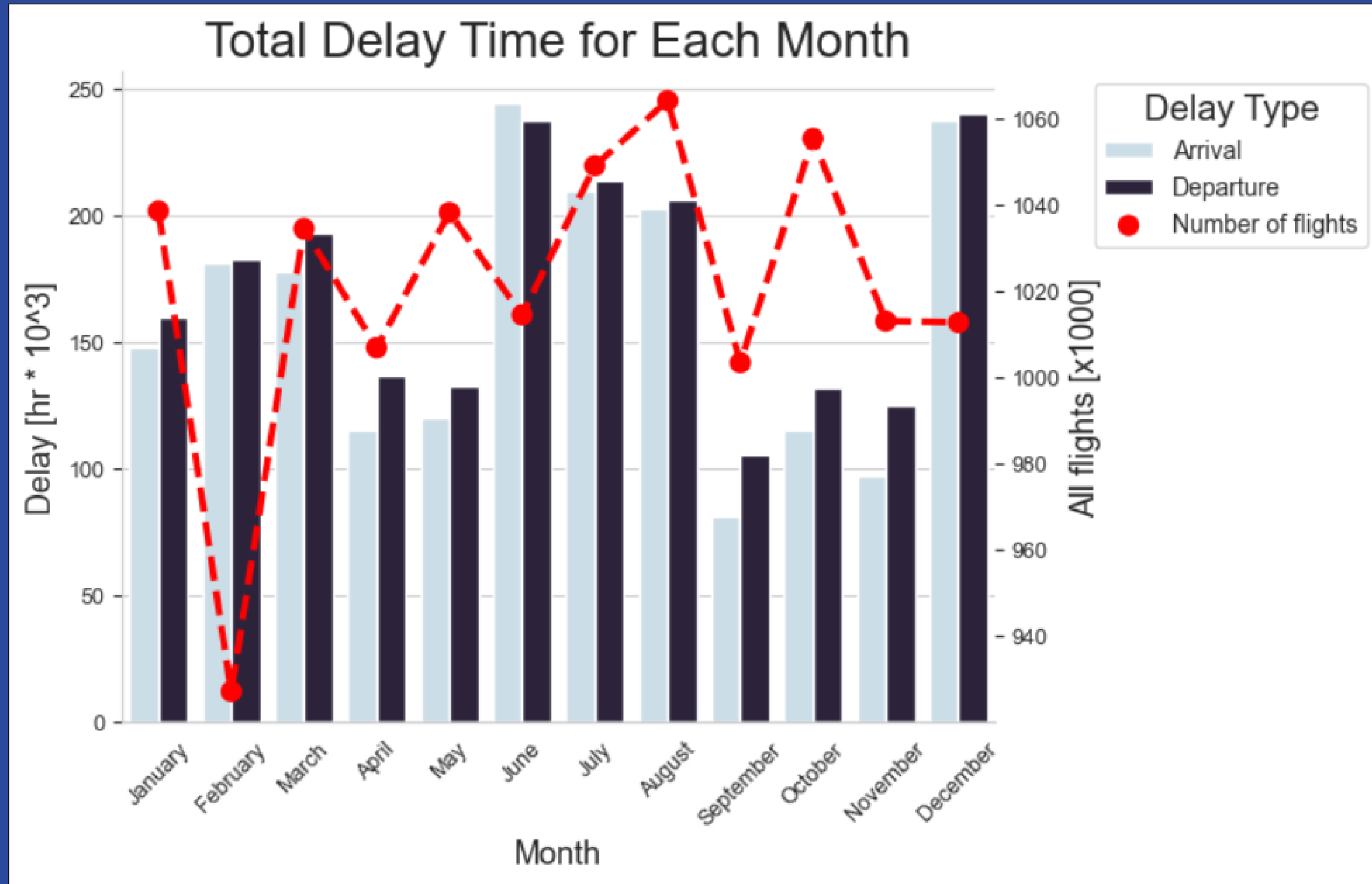
# CANCELLATIONS & DELAYS



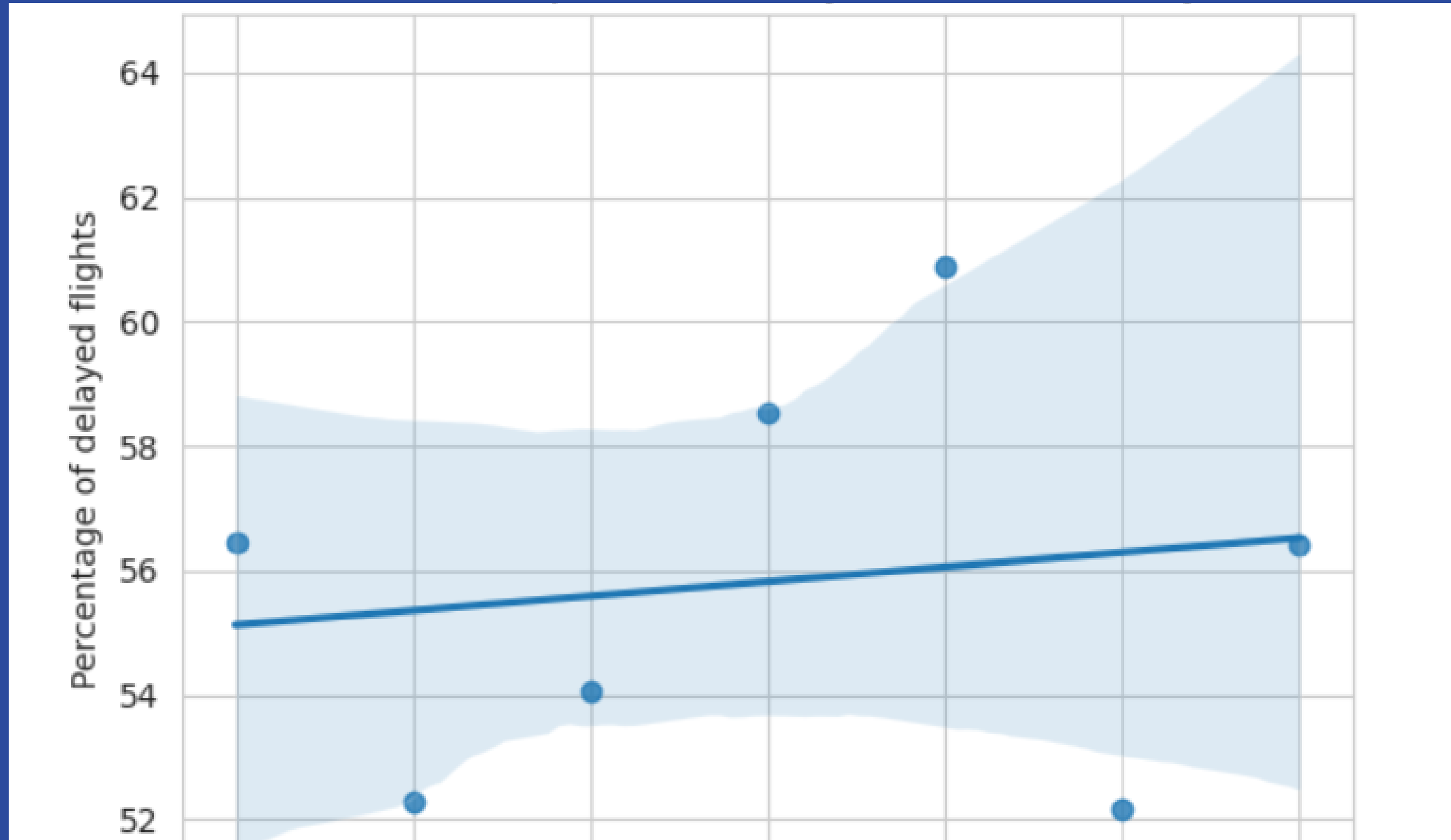
# CANCELLATIONS & DELAYS



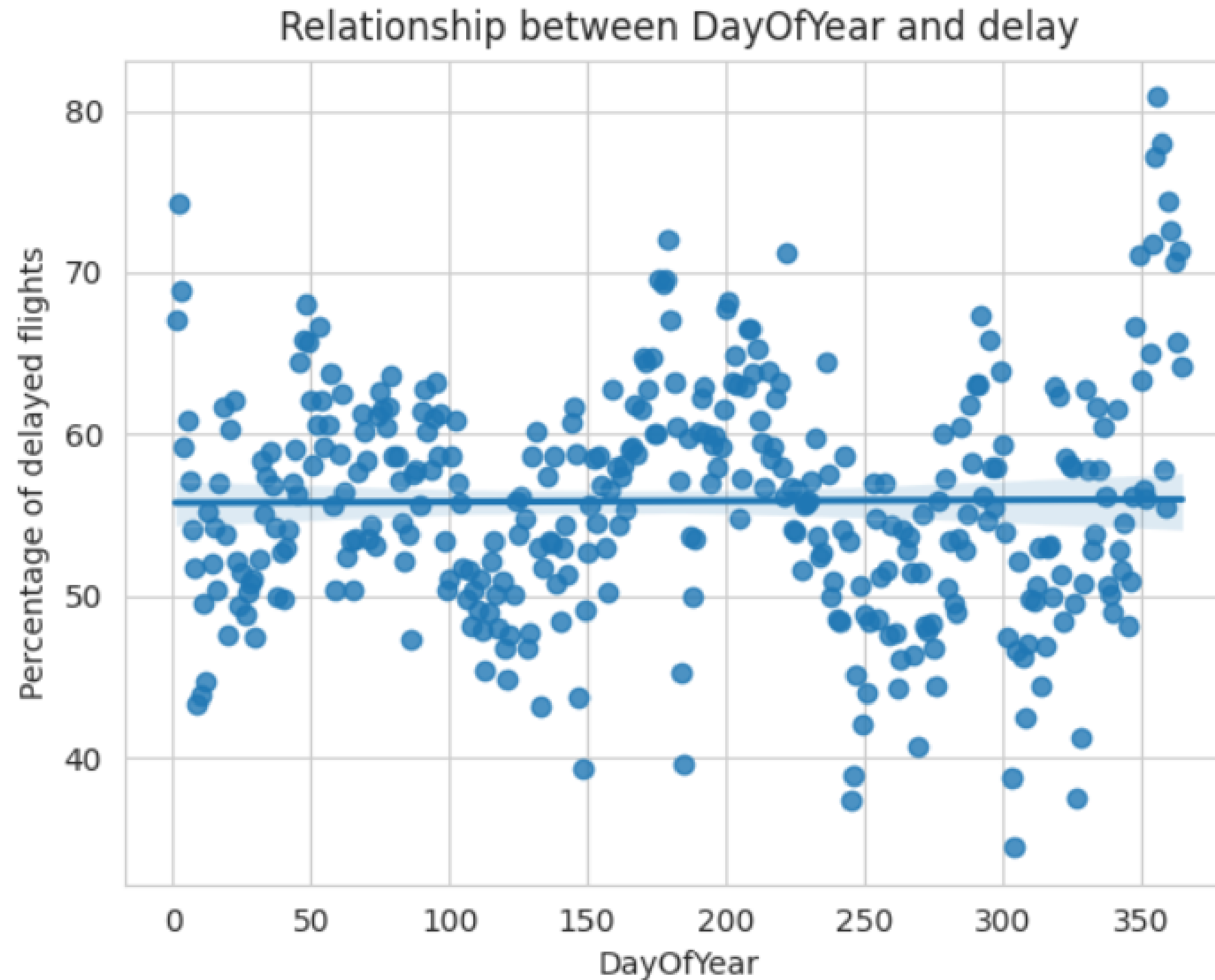
# CANCELLATIONS & DELAYS



# CANCELLATIONS & DELAYS

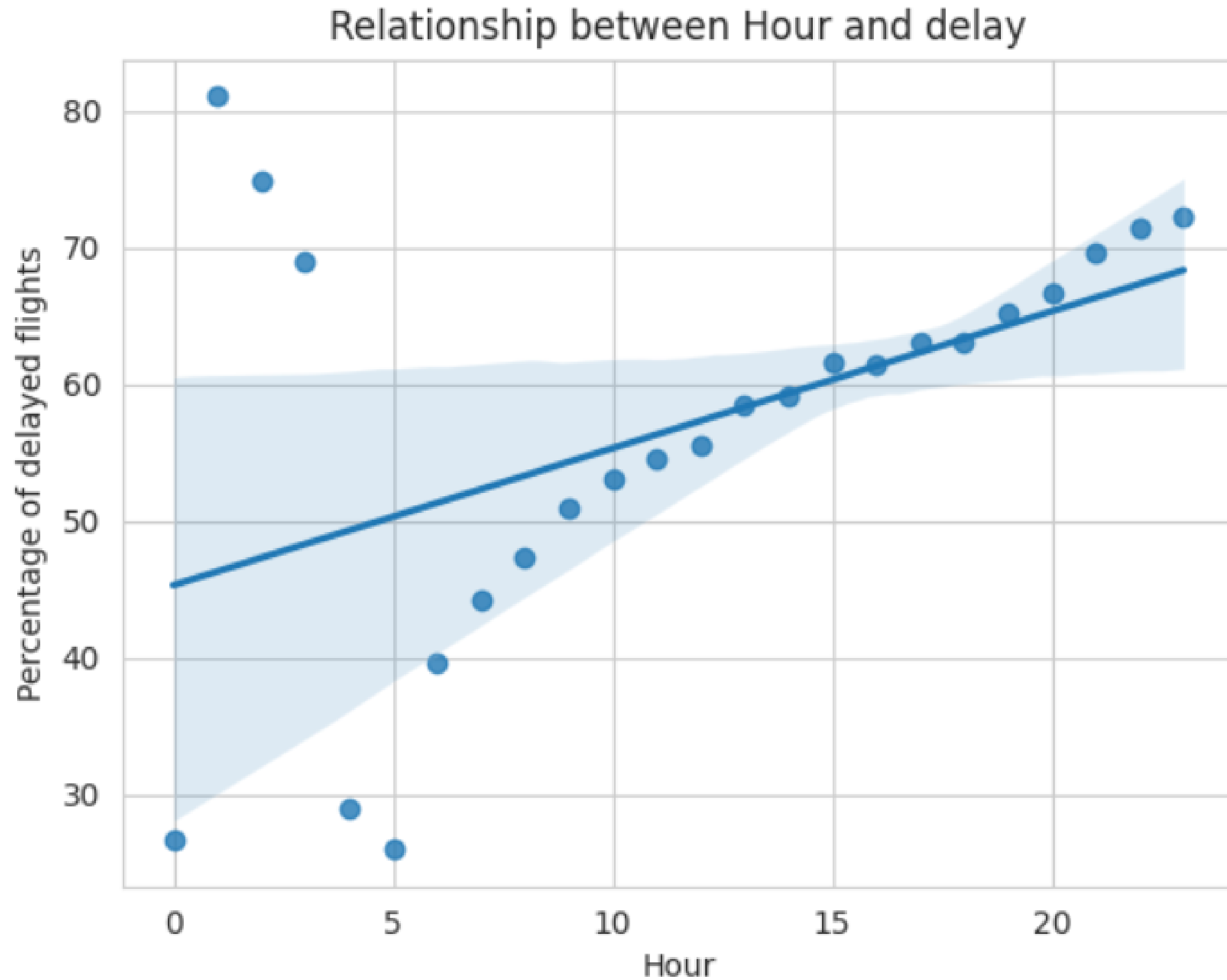


# CANCELLATIONS & DELAYS

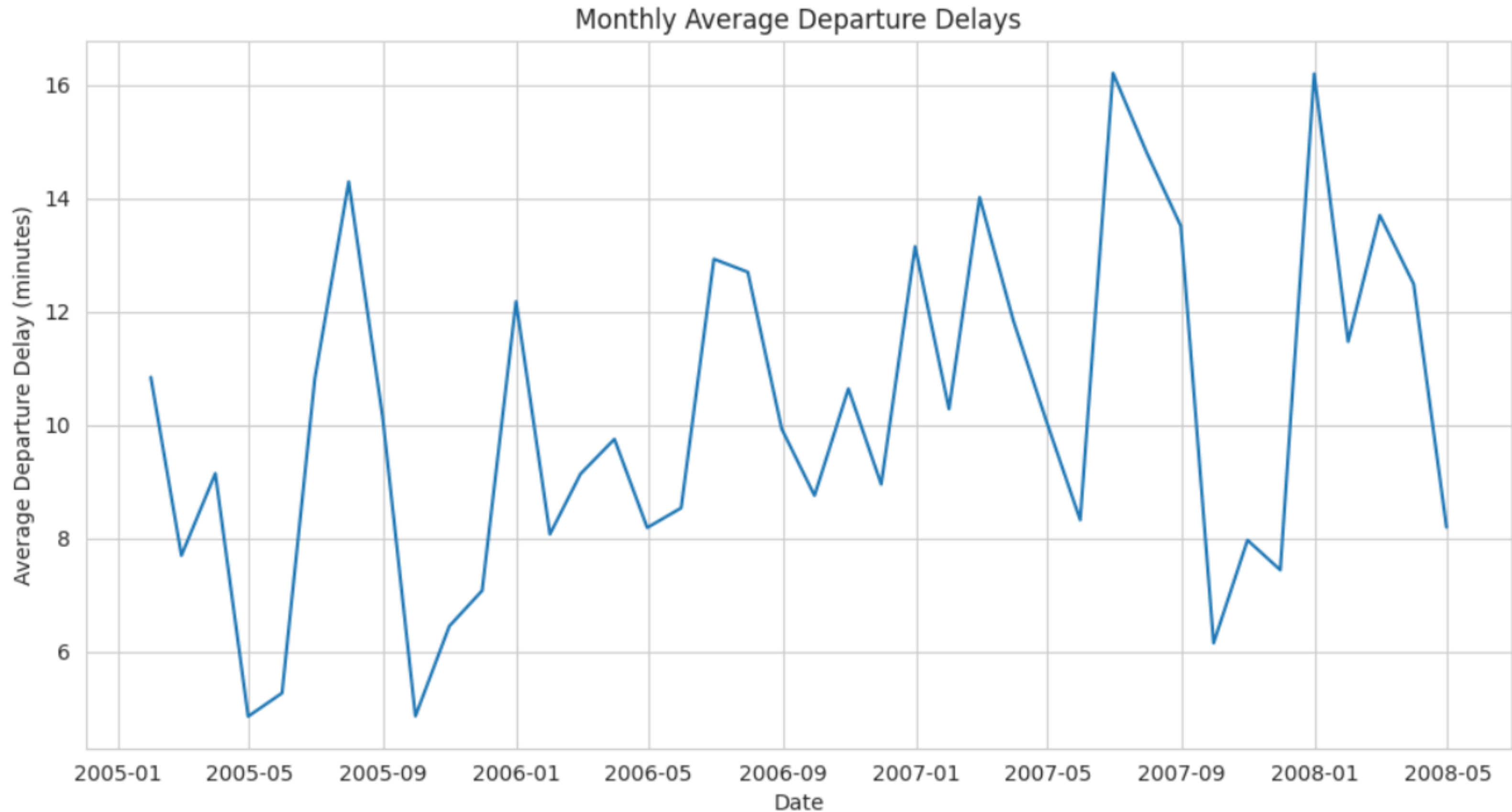




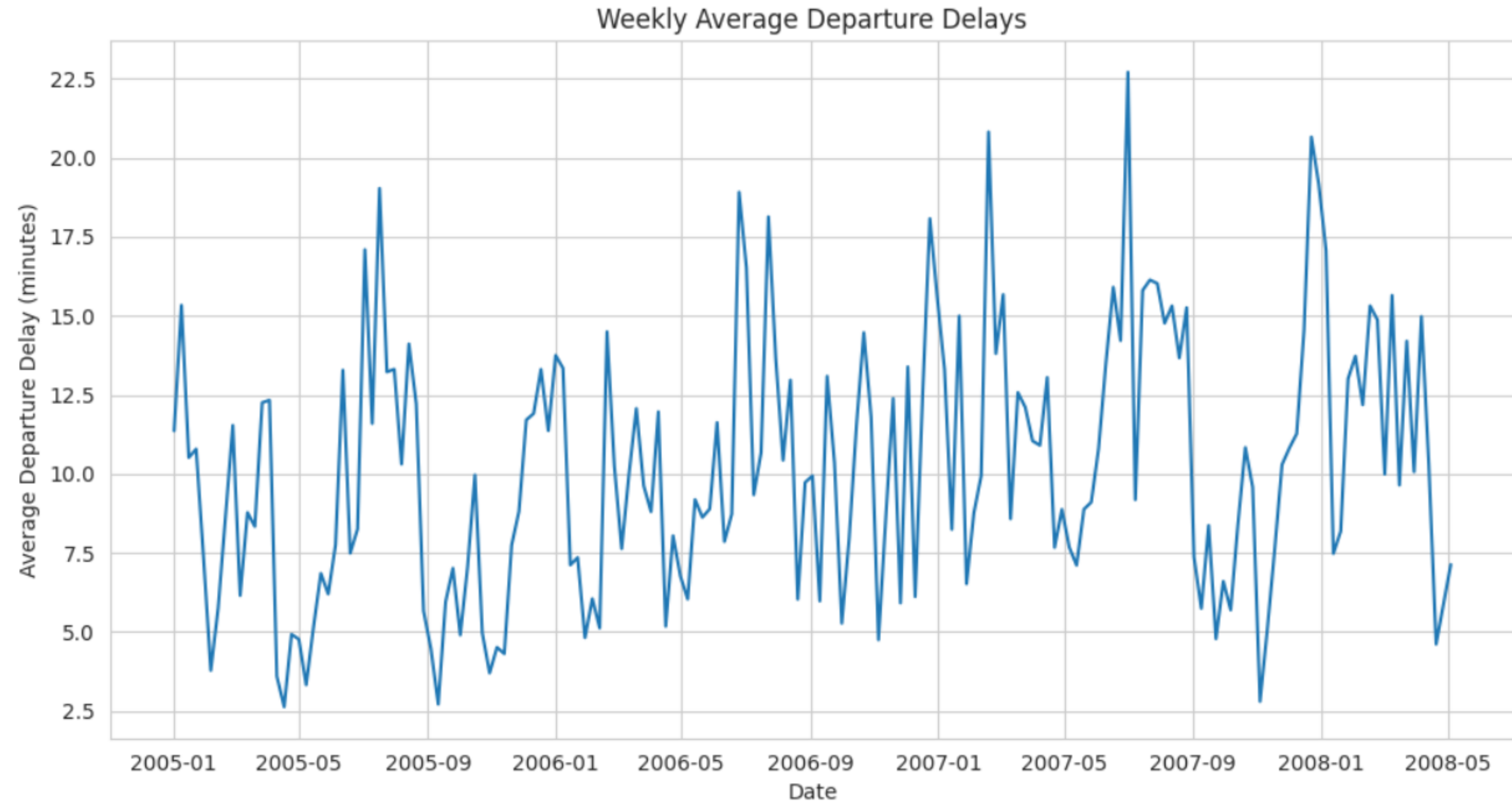
# CANCELLATIONS & DELAYS



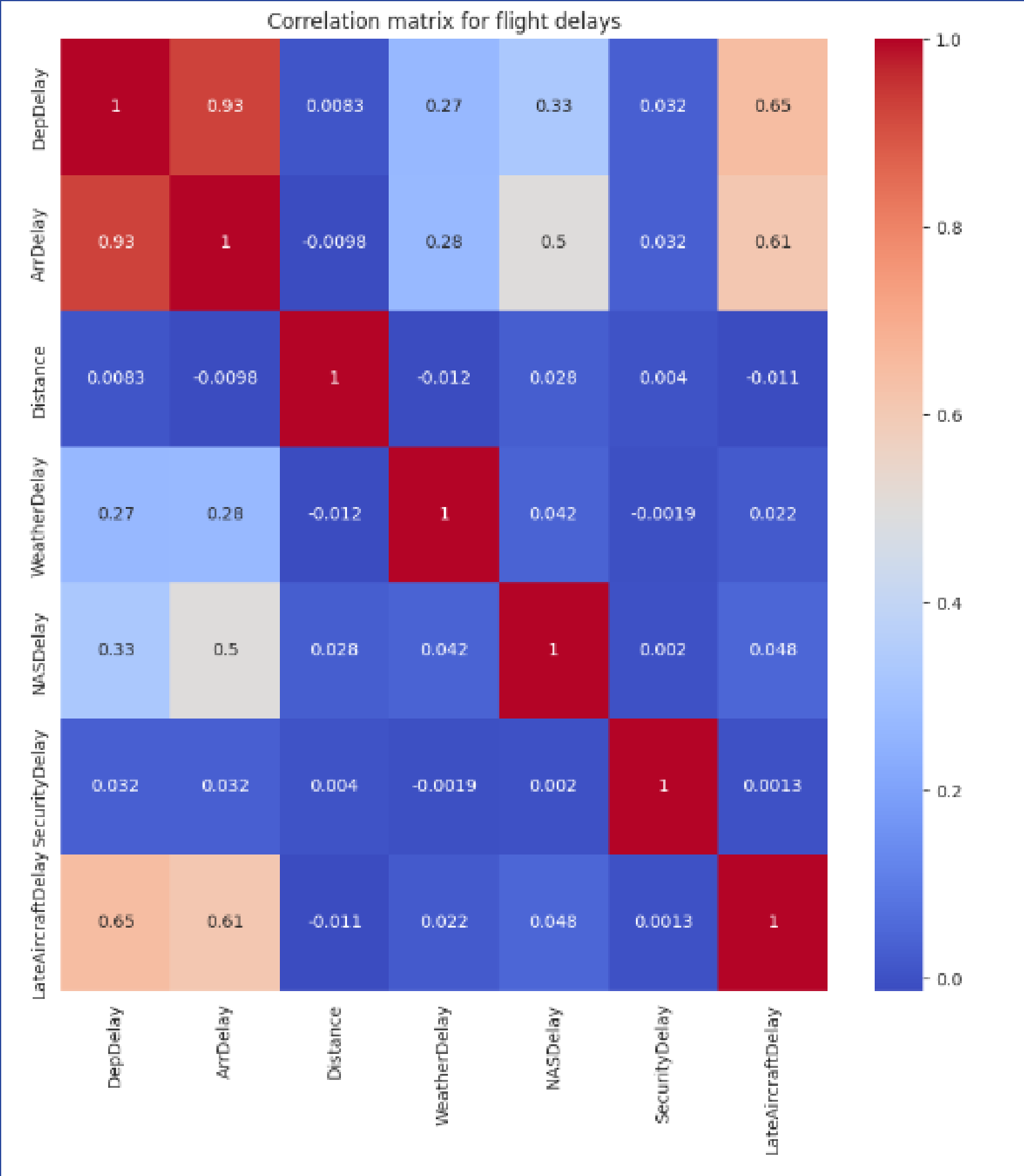
# CANCELLATIONS & DELAYS



# CANCELLATIONS & DELAYS



# CANCELLATIONS & DELAYS

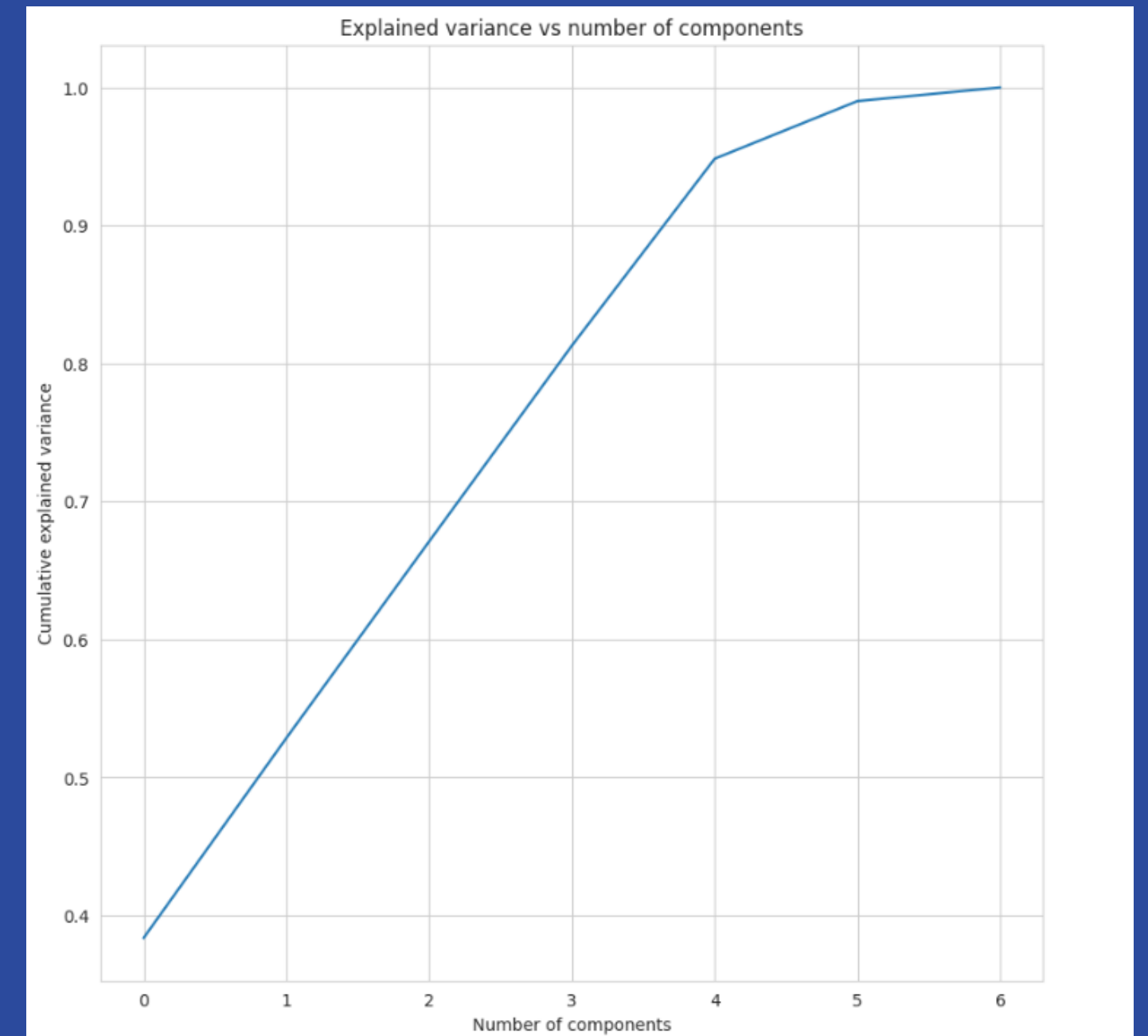
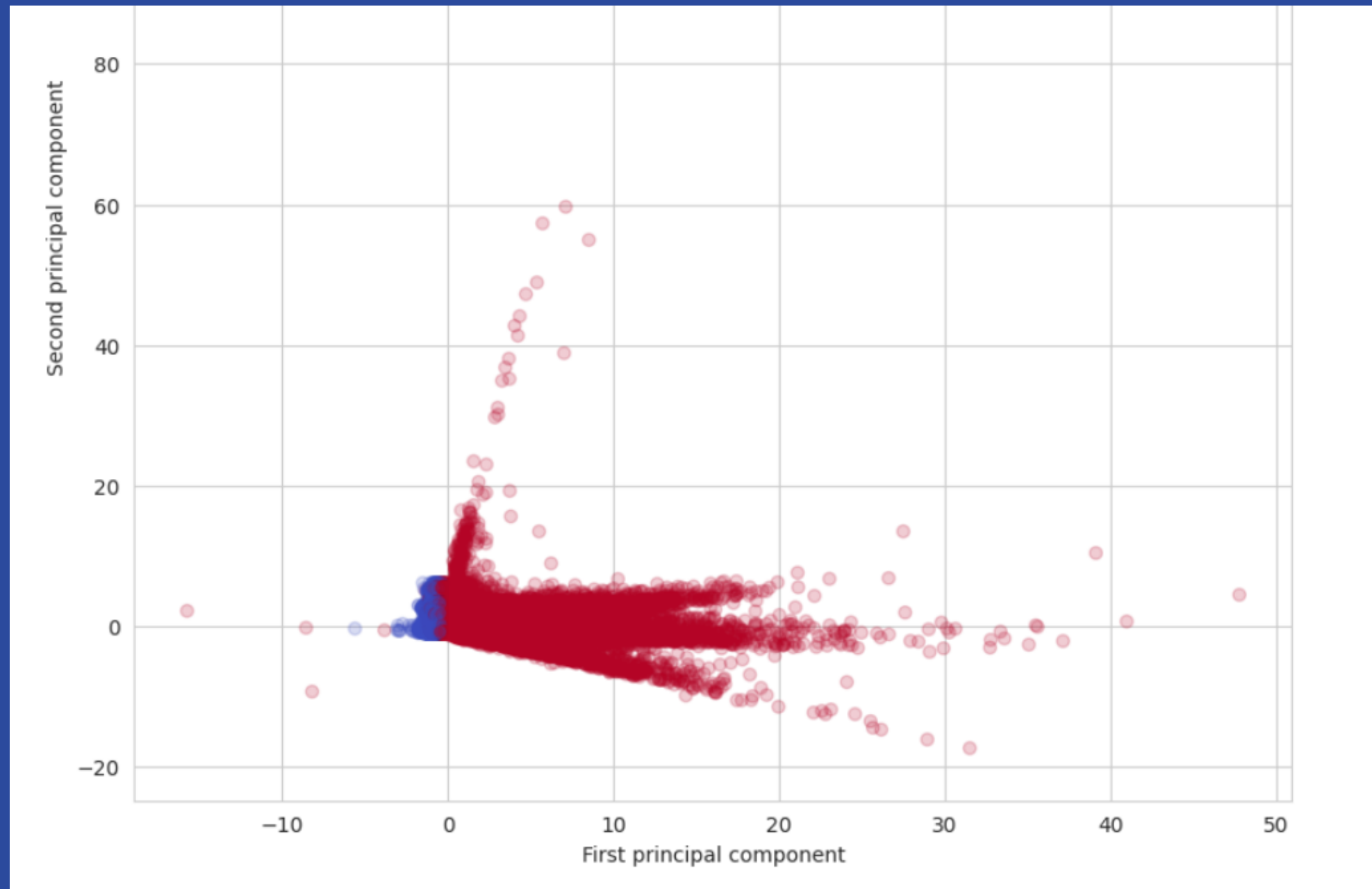


# AIRCRAFT TYPES VS. DELAYS

1	aircraft_type	engine_type	mean_CarrierDelay	percentage_delayed_by_CarrierDelay	mean_WeatherDelay	percentage_delayed_by_WeatherDelay
2	Fixed Wing Multi-Engine	Turbo-Prop	5.52	79.2	0.67	0.83
3	Rotorcraft	Turbo-Shaft	4.78	71.7	1.1	1.24
4	Fixed Wing Multi-Engine	Turbo-Jet	4.31	71.19	0.71	0.83
5	Fixed Wing Single-Engine	Turbo-Prop	3.9	70.71	1.15	1.32
6	Fixed Wing Multi-Engine	Turbo-Fan	4.1	70.62	0.89	1.01
7	Fixed Wing Single-Engine	4 Cycle	4.66	69.93	0.83	0.99
8	Fixed Wing Single-Engine	Reciprocating	4.32	69.18	1.43	1.64
9	Fixed Wing Multi-Engine	Reciprocating	4.44	68.09	1.06	1.16
10	Rotorcraft	4 Cycle	1.0	0.0	1.6	2.38

mean_NASDelay	percentage_delayed_by_NASDelay	mean_SecurityDelay	percentage_delayed_by_SecurityDelay	mean_LateAircraftDelay	percentage_delayed_by_LateAircraftDelay
1.85	2.56	0.03	0.08	3.71	6.28
5.76	9.79	0.03	0.07	6.12	7.64
4.65	9.25	0.03	0.07	5.28	8.1
5.37	8.4	0.01	0.03	7.01	9.14
4.43	7.43	0.03	0.07	5.84	8.87
5.35	10.36	0.02	0.03	5.71	8.07
5.66	9.71	0.01	0.02	6.27	8.53
5.37	8.92	0.02	0.05	6.67	8.76
7.5	2.38	0.0	0.0	26.2	19.05

# PCA



# PCA

Most important features according to first principal component:

	PC1	PC2	PC3	PC4	PC5	PC6	
DepDelay	0.575106	-0.042715	0.027486	-0.050920	-0.086034	0.494270	\
ArrDelay	0.588922	0.026583	-0.011336	0.014772	0.067602	0.300717	
Distance	0.000355	0.851265	-0.132267	-0.159719	-0.481799	-0.000900	
WeatherDelay	0.198766	-0.139079	-0.214907	0.790486	-0.448910	-0.260664	
NASDelay	0.300755	0.410022	-0.182410	0.213455	0.699586	-0.380250	
SecurityDelay	0.022410	0.203990	0.937249	0.279558	0.010423	-0.034320	
LateAircraftDelay	0.438132	-0.209150	0.154037	-0.472329	-0.254654	-0.671976	

	PC7
DepDelay	-0.642165
ArrDelay	0.746402
Distance	0.014554
WeatherDelay	-0.025108
NASDelay	-0.169061
SecurityDelay	-0.003362
LateAircraftDelay	-0.032760

**Percentage of delayed flights by airline**

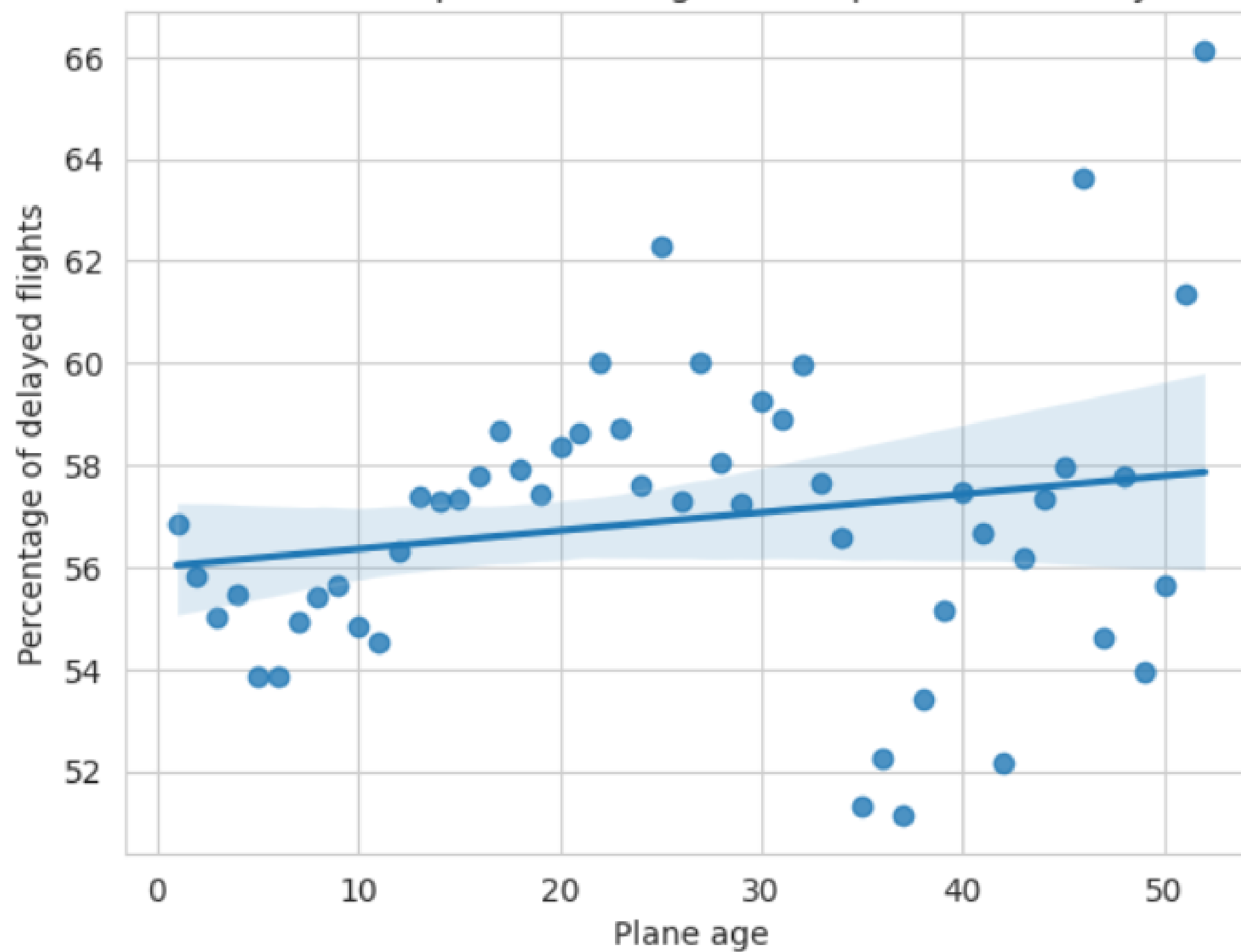
**Most often delayed routes**

**Routes outlying average**

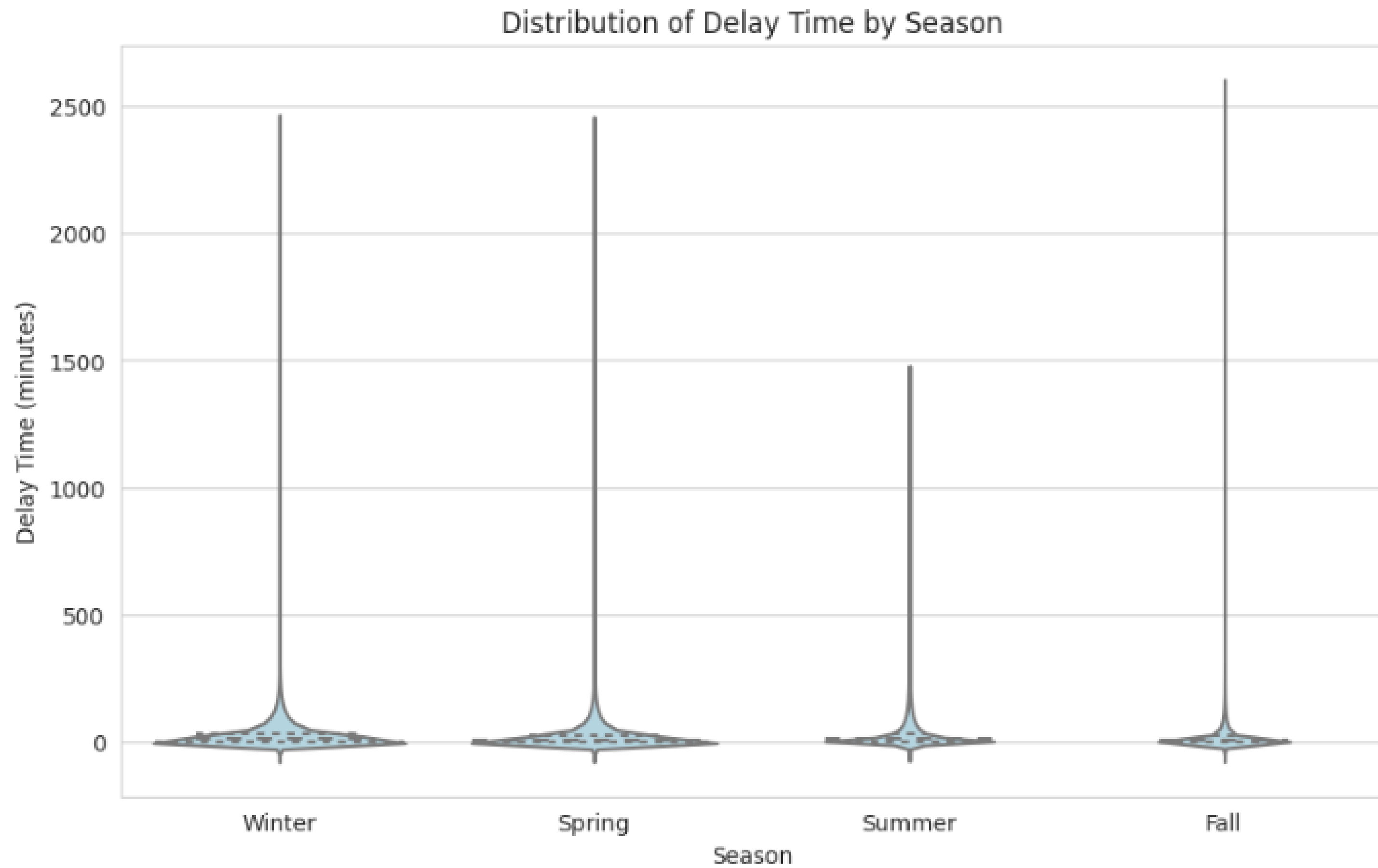


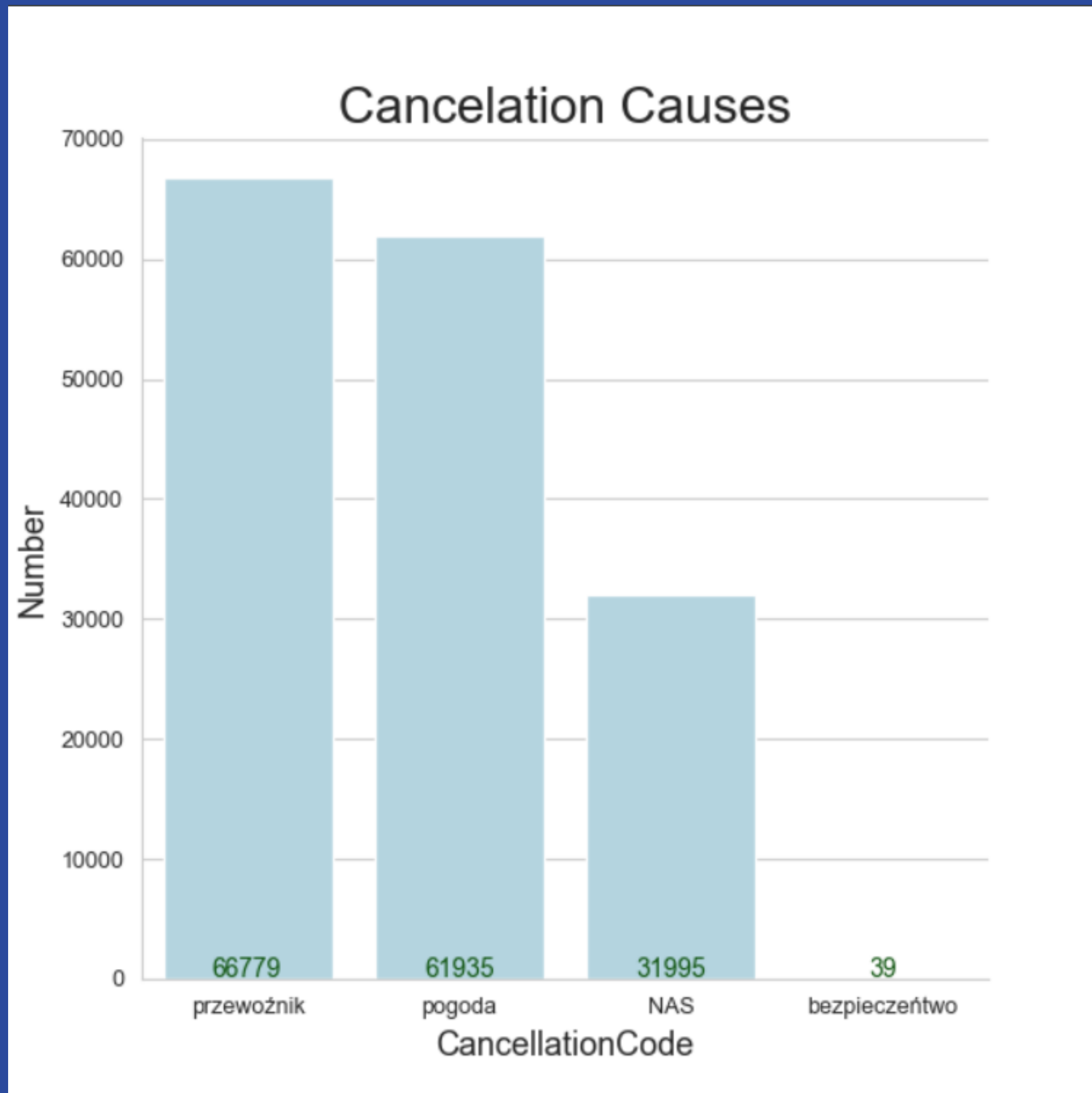
# WHY? & HOW TO SOLVE?

Relationship between age of the plane and delay

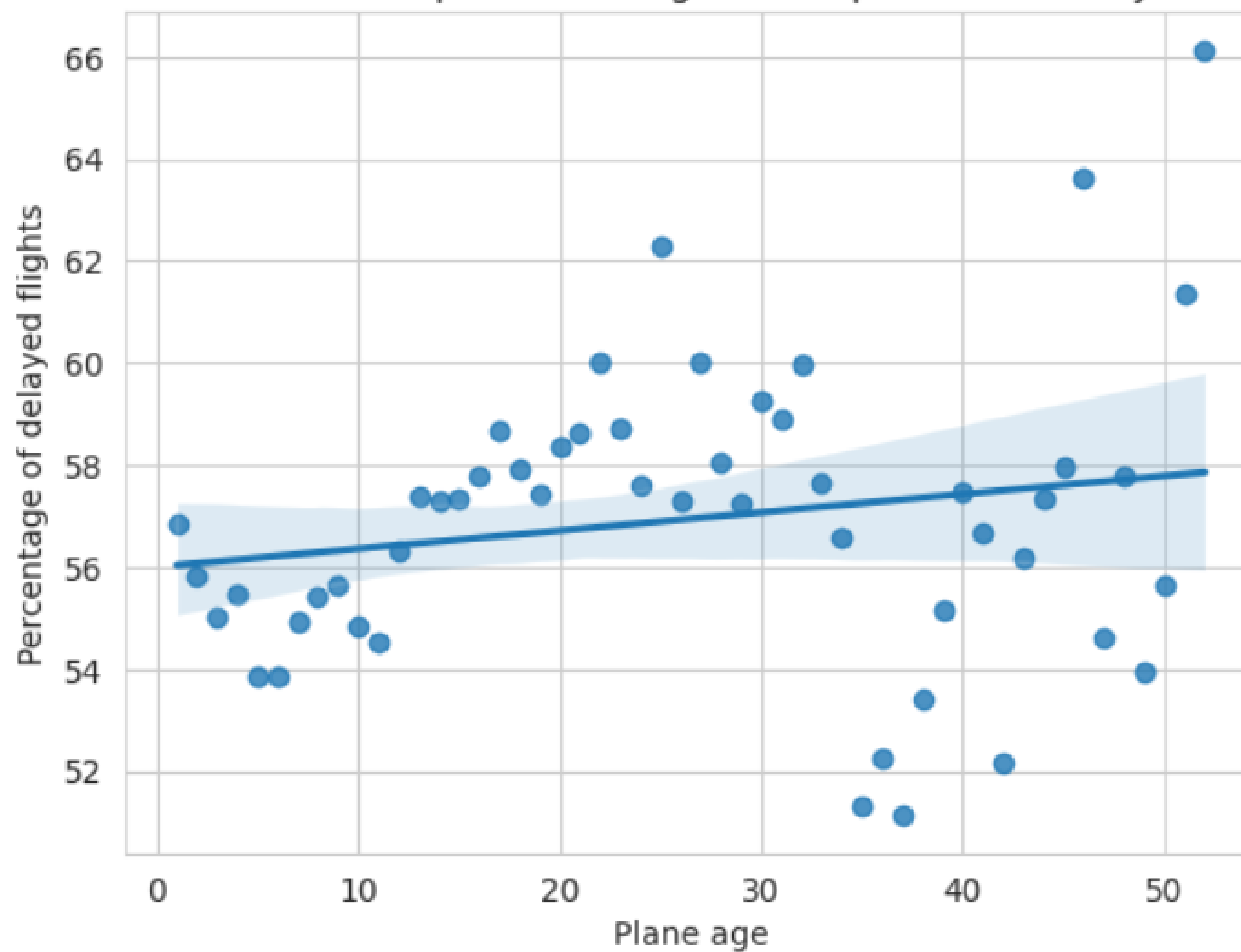


Average delay time vs distance			
	mean_dep_delay	mean_arr_delay	mean_delay
DistanceGroup			
3001-3500	13.027144	9.241684	11.134415
501-1000	11.568981	10.354849	10.961915
1001-1500	11.176467	9.192326	10.184397
1501-2000	11.190056	9.102136	10.146095
0-500	10.211769	9.726192	9.968981
2001-2500	10.796847	8.095785	9.446316
2501-3000	10.041550	7.623671	8.832610
3501-4000	8.665770	7.519119	8.092444
4501-5000	9.660941	6.135066	7.898004
4001-4500	7.255764	1.506122	4.380943
5001-5500	NaN	NaN	NaN
5501-6000	NaN	NaN	NaN
6001-6500	NaN	NaN	NaN
6501-7000	NaN	NaN	NaN
7001-7500	NaN	NaN	NaN
7501-8000	NaN	NaN	NaN
8001-8500	NaN	NaN	NaN
8501-9000	NaN	NaN	NaN
9001-9500	NaN	NaN	NaN
9501-10000	NaN	NaN	NaN
10001+	NaN	NaN	NaN





Relationship between age of the plane and delay



# Most important airports

Most important airports according to PageRank centrality:

PHX 0.06157768311578246  
LAS 0.04973906848057322  
CVG 0.046670936869472796  
DEN 0.03564864160625193  
SLC 0.032081069450617464  
ORD 0.031624280701629585  
CLT 0.025561061157029084  
MDW 0.022593954993377935  
PHL 0.02109845289723674  
LAX 0.020060513600105093

Most important airports according to number of flights on the routes:

PHX 0.15087412587949986  
CLT 0.08678457266581356  
LAS 0.0735872215872303  
PHL 0.05073018206140891  
DCA 0.03881532679760962  
BOS 0.0268880241603067  
PIT 0.02197702209503685  
LAX 0.021274756980257936  
BWI 0.01806170706515466  
LGA 0.017760167358634206

# Delay airports

	DepDelay	ArrDelay	MeanDelay
Location			
Chicago	12.381458	11.429193	11.905325
New_York	12.005730	11.042230	11.523979
St_Thomas	9.895971	9.666268	9.781120
Anchorage	9.069686	7.525334	8.297510
Los_Angeles	8.760744	7.080968	7.920856
Denver	8.190011	7.641464	7.915738
Phoenix	9.351537	6.474339	7.912938
Puerto_Rico	7.796348	6.594069	7.195209
Honolulu	-0.016226	0.211398	0.097586
Adak	1.924528	-16.126213	-7.100842

	Delayed
OriginLoc	
Phoenix	57.218871
St_Thomas	56.956933
Chicago	56.861763
New_York	56.852268
Los_Angeles	55.604763
Denver	53.810278
Puerto_Rico	51.243662
Anchorage	50.401391
Adak	42.477876
Honolulu	33.769711

	DepDelay
OriginLoc	
Chicago	12.381458
New_York	12.005730
St_Thomas	9.895971
Phoenix	9.351537
Anchorage	9.069686
Los_Angeles	8.760744
Denver	8.190011
Puerto_Rico	7.796348
Adak	1.924528
Honolulu	-0.016226



# Delay - aircrafts

		mean_dep_delay	mean_arr_delay	mean_delay
aircraft_type	engine_type			
Rotorcraft	Turbo-Shaft	13.151904	12.374250	12.763077
Fixed Wing Single-Engine	Turbo-Prop	12.922153	12.173966	12.548059
	Reciprocating	12.653880	12.072250	12.363066
Fixed Wing Multi-Engine	Reciprocating	12.626246	11.499735	12.062990
Fixed Wing Single-Engine	4 Cycle	12.216573	11.295675	11.756124
Fixed Wing Multi-Engine	Turbo-Fan	11.227782	9.790485	10.509134
	Turbo-Jet	10.472041	9.937351	10.204697
	Turbo-Prop	7.426540	7.542241	7.484391
Rotorcraft	4 Cycle	7.904762	3.547619	5.726191

		delayed_percentage	total_flights
aircraft_type	engine_type		
Fixed Wing Single-Engine	4 Cycle	58.967675	7672
Fixed Wing Multi-Engine	Turbo-Jet	57.340006	4134438
	Reciprocating	56.284261	15141
Fixed Wing Single-Engine	Turbo-Prop	56.061538	3250
Rotorcraft	Turbo-Shaft	55.978742	4516
Fixed Wing Single-Engine	Reciprocating	55.749696	43611
Fixed Wing Multi-Engine	Turbo-Fan	55.668457	8842406
Rotorcraft	4 Cycle	52.380952	42
Fixed Wing Multi-Engine	Turbo-Prop	48.413268	508183

# PREDICTIONS

```
Mean squared error: 226.48902287689813
Mean absolute error: 9.447113426282387
Root mean squared error: 15.049552248385934
R2 score: 0.8311590999314569
```

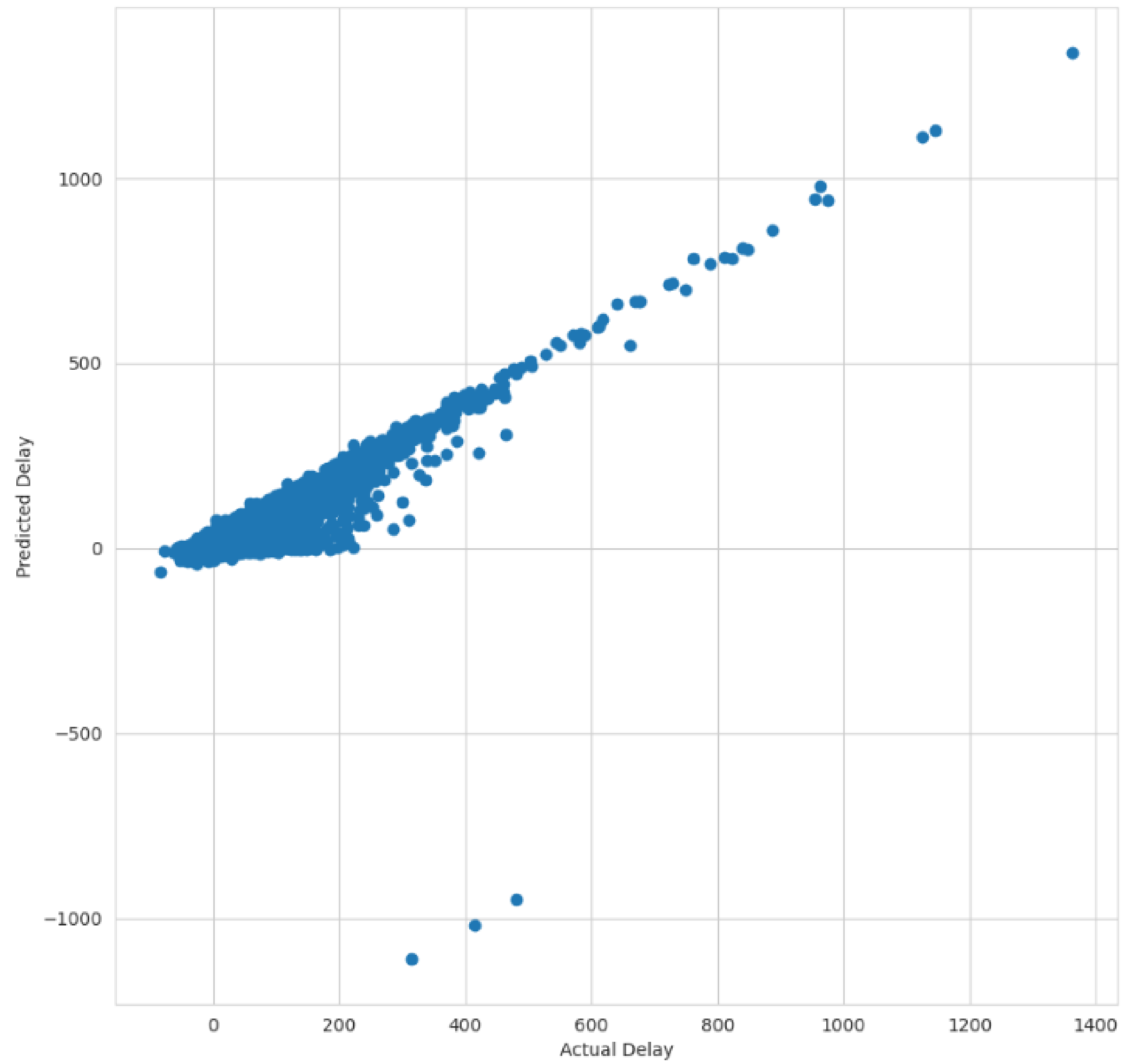
```
Random flight:
UniqueCarrier      UA
Origin             SFO
Dest               ORD
Departure          2006-01-08 10:42:00
Arrival            2006-01-08 16:42:00
ArrDelay           -9.0
Name: 174683, dtype: object
Predicted delay: -6.574951171875
Actual delay: -6.0
```

```
Random flight:
UniqueCarrier      UA
Origin             FLL
Dest               ORD
Departure          2006-01-23 06:13:00
Arrival            2006-01-23 08:15:00
ArrDelay           -14.0
Name: 185306, dtype: object
Predicted delay: -5.183349609375
Actual delay: -19.0
```

```
Random flight:
UniqueCarrier      MQ
Origin             BOS
Dest               BWI
Departure          2006-01-01 13:55:00
Arrival            2006-01-01 15:24:00
ArrDelay           -4.0
Name: 277926, dtype: object
Predicted delay: -2.659423828125
Actual delay: -1.0
```

```
Random flight:
UniqueCarrier      AA
Origin             BOS
Dest               ORD
Departure          2006-01-27 05:43:00
Arrival            2006-01-27 07:14:00
ArrDelay           -23.0
Name: 336317, dtype: object
Predicted delay: -3.466552734375
Actual delay: -12.0
```

```
Random flight:
UniqueCarrier      UA
Origin             SMF
Dest               DEN
Departure          2006-01-23 13:53:00
Arrival            2006-01-23 17:17:00
ArrDelay           -3.0
Name: 158881, dtype: object
Predicted delay: 1.46240234375
Actual delay: -4.0
```





# Analyze, solve, predict!

PAWEŁ POZORSKI    IGOR KOŁODZIEJ    NAZARII BIHNIAK