

iscte

SINTRA
TECNOLOGIAS DIGITAIS
ECONOMIA E SOCIEDADE

ANÁLISE DE PADRÕES DE MOBILIDADE AÉREA

Vasco Silva
Guilherme Correia



Funcionamento do sistema

<input type="checkbox"/>	Name ↑	Container ID	Image	Port(s)	CPU (%)	Last started	Actions
<input type="checkbox"/>	projecto-big-data	-	-	-	7.16%	8 hours ago	
<input type="checkbox"/>	jupyter-1	759d3dd52203	projecto-big-data-jupyter	8888:8888	6.92%	8 hours ago	
<input type="checkbox"/>	mongodb-1	ec36cc6cfe08	projecto-big-data-mongodb	27017:27017	0.24%	8 hours ago	

Funções do sistema:

- Container de docker que abre instâncias para MongoDB e Jupyter com todas as bibliotecas necessárias para ligações entre eles
- No MongoDB armazenamos os nossos datasets para análise em várias coleções

projeto-big-data

admin

config

local

metro_lisboa

projeto

atrasos

batch_aeroporto_resumo

batch_rendimento_resumo

streaming_aeroporto_resumo

streaming_rendimento_resumo

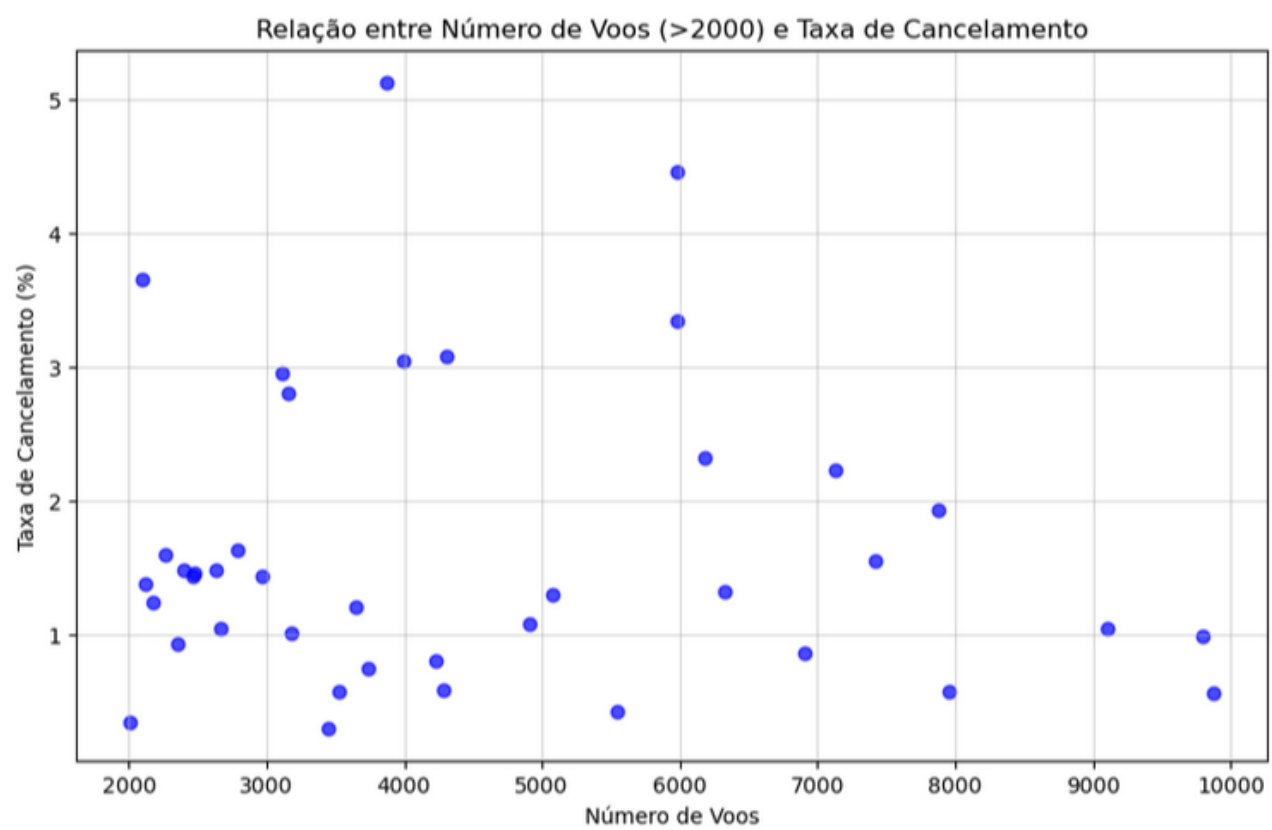
voos

Funcionamento do sistema

- Em Jupyter é utilizado o PySpark para fazer a ligação com MongoDB e consequentemente a análise dos dados
- Por último temos um script para processamento em batch e streaming de dados, ligado ao MongoDB

```
df_2000 = df_comb_voos_canc_pd[df_comb_voos_canc_pd["num_flights"] > 2000]

plt.figure(figsize=(10, 6))
plt.scatter(df_2000["num_flights"], df_2000["cancel_rate"], alpha=0.7, color="blue")
plt.title("Relação entre Número de Voos (>2000) e Taxa de Cancelamento")
plt.xlabel("Número de Voos")
plt.ylabel("Taxa de Cancelamento (%)")
plt.grid(alpha=0.5)
plt.show()
```



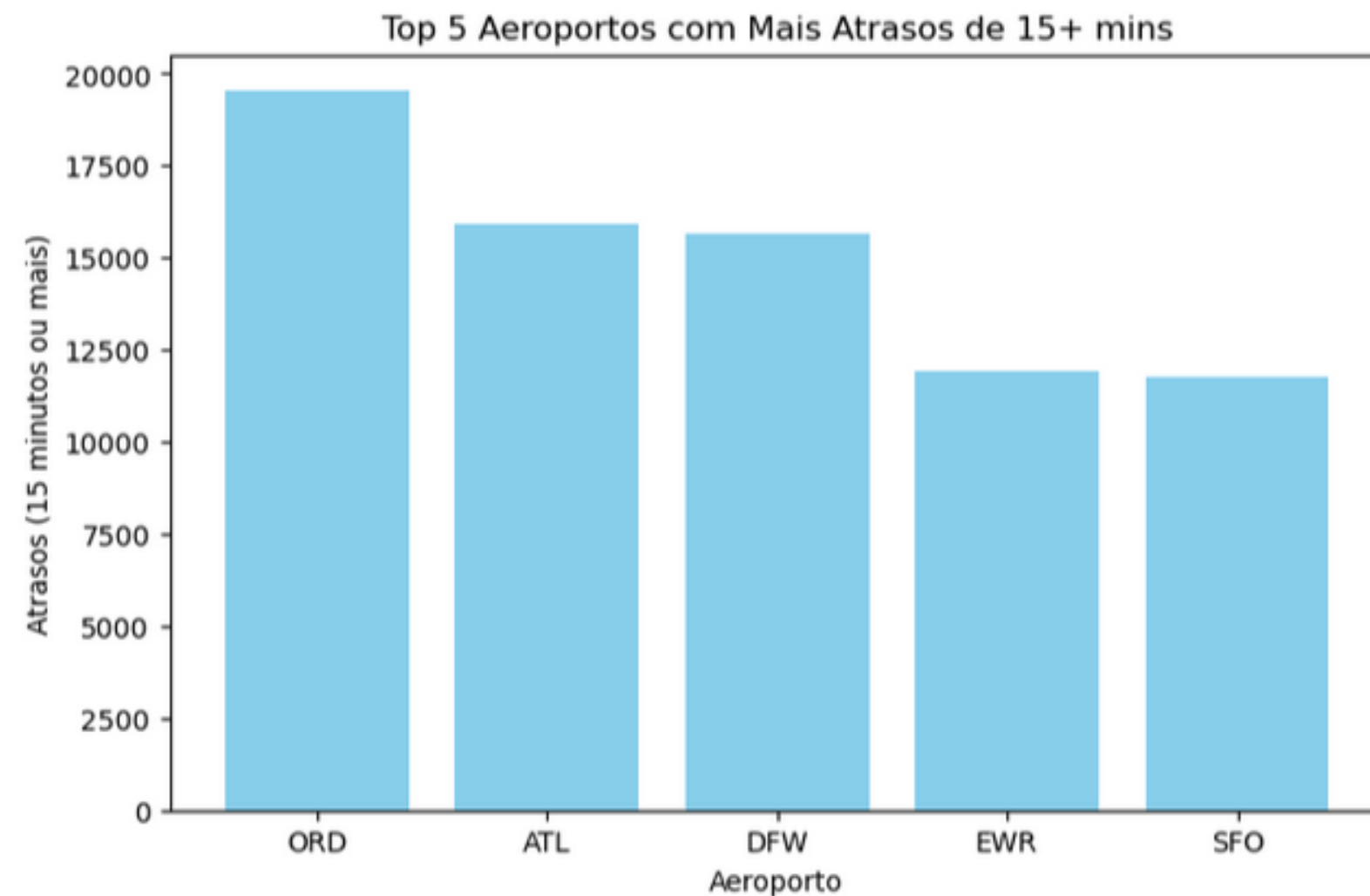
Resumo Incremental dos Aeroportos:

airport	airport_name	total_flights	total_delayed_flights	avg_delay_minutes
HVN	New Haven, CT: Tw...	256.0	49.0	1146.3333333333333
BLV	Belleville, IL: S...	277.0	69.0	1595.6666666666667
LWS	Lewiston, ID: Lew...	178.0	13.0	246.66666666666666
TVC	Traverse City, MI...	1568.0	237.0	842.7777777777778
DRO	Durango, CO: Dura...	806.0	172.0	1441.625
IND	Indianapolis, IN:...	12338.0	2581.0	4116.3333333333333
HYA	Hyannis, MA: Cape...	66.0	14.0	308.33333333333333
ACK	Nantucket, MA: Na...	723.0	136.0	1368.0
FCA	Kalispell, MT: Gl...	1062.0	157.0	793.0
TUL	Tulsa, OK: Tulsa ...	4439.0	941.0	1647.1842105263158
SEA	Seattle, WA: Seat...	39485.0	8870.0	16606.48275862069
HRL	Harlingen/San Ben...	844.0	148.0	602.3333333333334
CPR	Casper, WY: Caspe...	280.0	14.0	174.0
PBG	Plattsburgh, NY: ...	210.0	28.0	258.25
HIB	Hibbing, MN: Rang...	156.0	6.0	101.66666666666667
ISN	Williston, ND: Sl...	397.0	50.0	716.3333333333334
GPT	Gulfport/Biloxi, ...	1127.0	274.0	1228.1176470588234
SAF	Santa Fe, NM: San...	350.0	69.0	998.3333333333334
STC	St. Cloud, MN: St...	17.0	4.0	129.0
PUB	Pueblo, CO: Puebl...	233.0	30.0	649.0

only showing top 20 rows

Resumo Incremental de Rendimentos por Companhia Aérea:

AirlineCompany	total_revenue	avg_ticket_price
UA	2.071318267000008E7	283.22902865165594
NK	2748692.030000001	107.09778058007555
AA	2.0905140830000065E7	273.0445403523927
B6	9681379.369999984	207.47527300760294
DL	2.5022957509999994E7	264.8980986183302
F9	2236193.5699999994	95.0715259646069
HA	2191704.0099999984	304.20208069858495
G4	2326032.21	98.497109401516
AS	9457367.040000007	248.36277990351047
SY	644862.7200000001	164.8247380952381
WN	2.693869883000006E7	191.3949029666878



Dest	AveragePrice
EWR	277.48
DFW	259.76
SFO	294.81
ATL	238.68
ORD	249.66

Análise de resultados

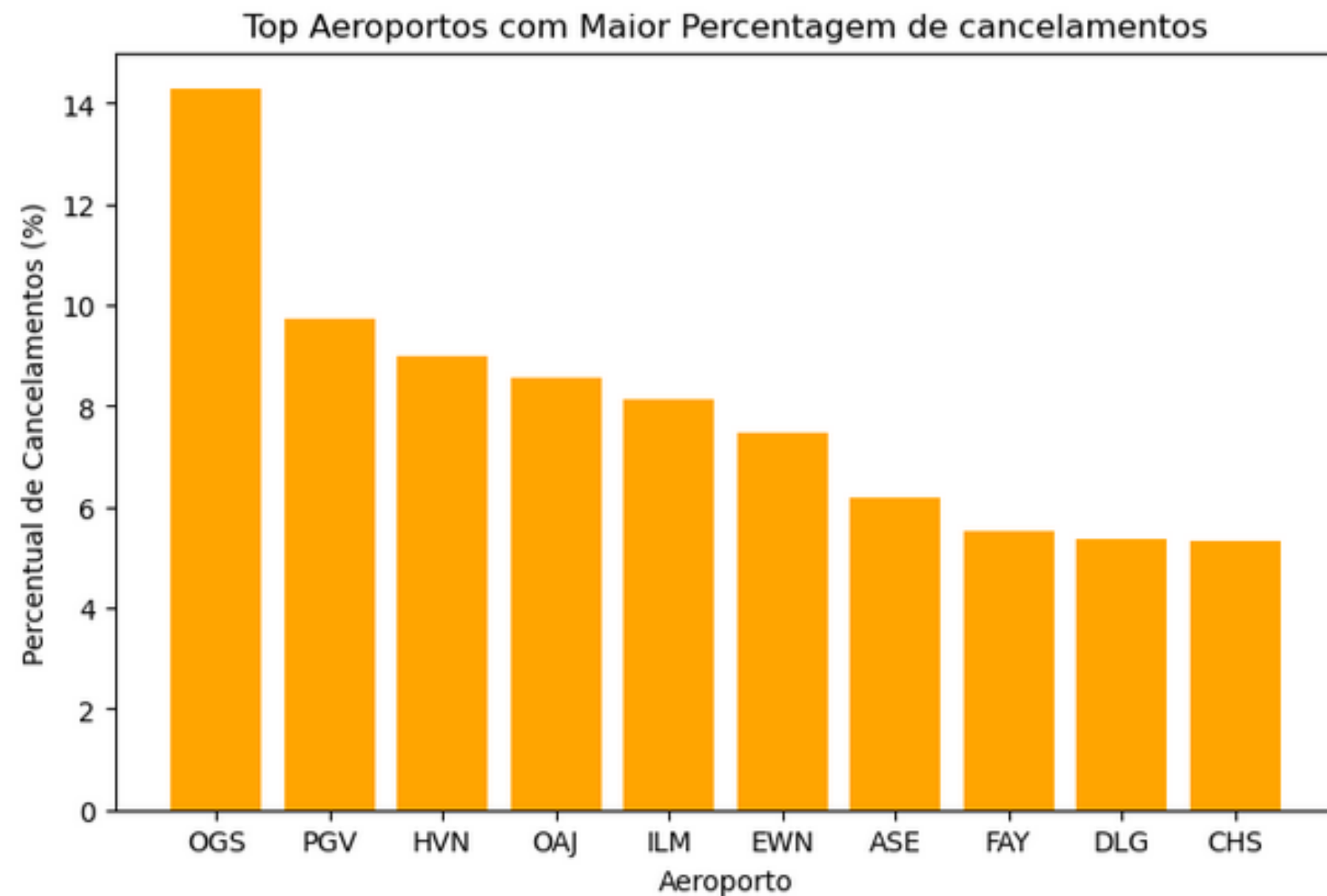
Os 5 aeroportos que mais atrasos significativos apresentam encontram-se em cidades importantes ou com afluência turística.

- ORD: Chicago
- ATL: Atlanta
- DFW: Dallas/Fort Worth
- EWR: Newark, New Jersey
- SFO: San Francisco

Análise de resultados

Os aeroportos com maior percentagem de cancelamentos estão localizados em áreas com condições climáticas adversas ou que possuem características específicas que influenciam os índices de cancelamento:

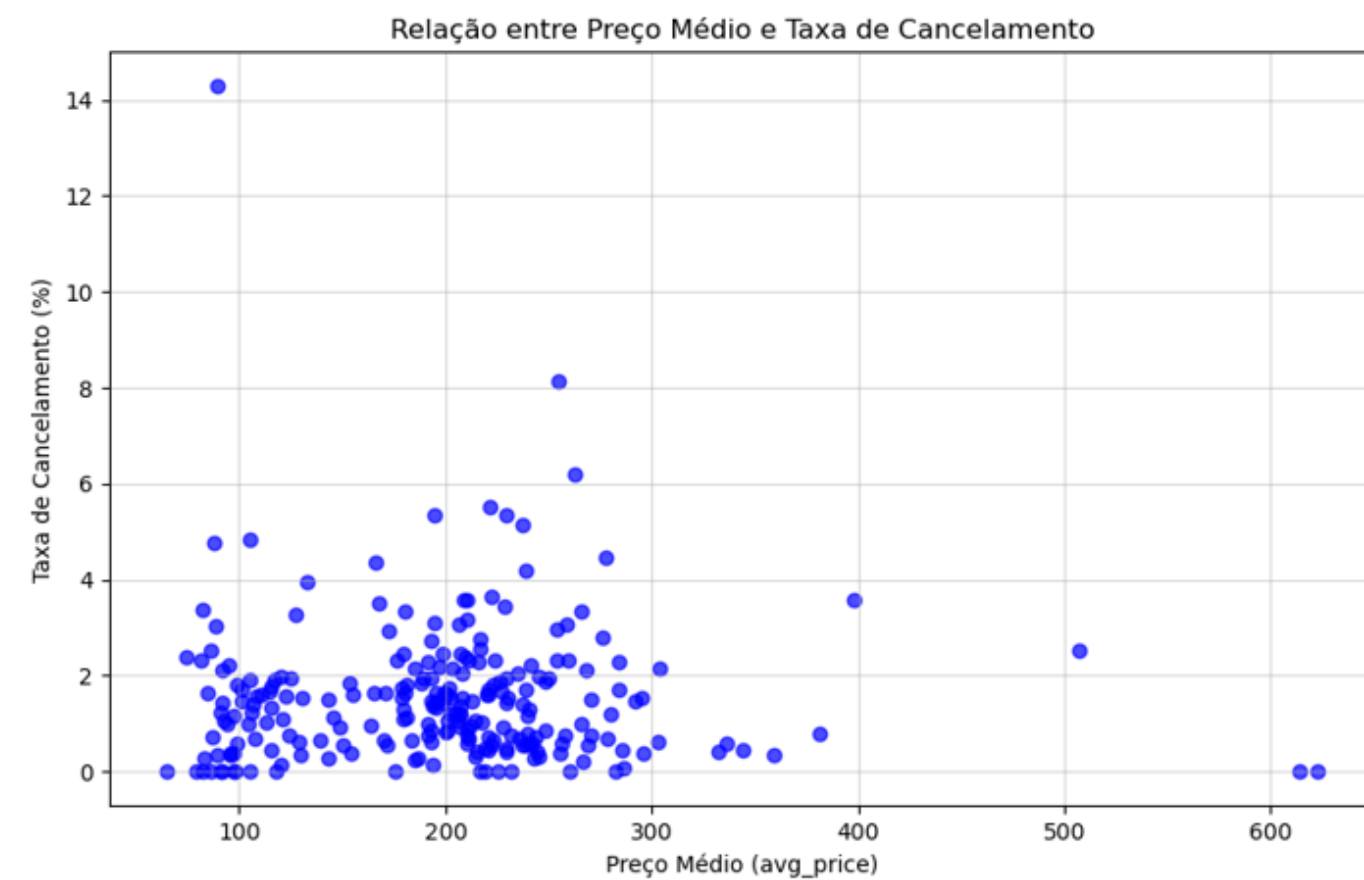
- OGS: Ogdensburg International Airport (Ogdensburg, Nova York)
- PGV: Pitt-Greenville Airport
- HVN: Tweed-New Haven Airport
- OAJ: Albert J. Ellis Airport



Relação entre Preço Médio e taxa de Cancelamento

- É possível perceber que a grande maioria dos bilhetes se mantém entre os 100 e os 300 dólares.
- Bilhetes de voos acima dos 300 têm taxas de cancelamento baixas, com exceções.
- OGS sendo o aeroporto com bilhetes mais baratos, tem uma taxa de cancelamento muito baixa

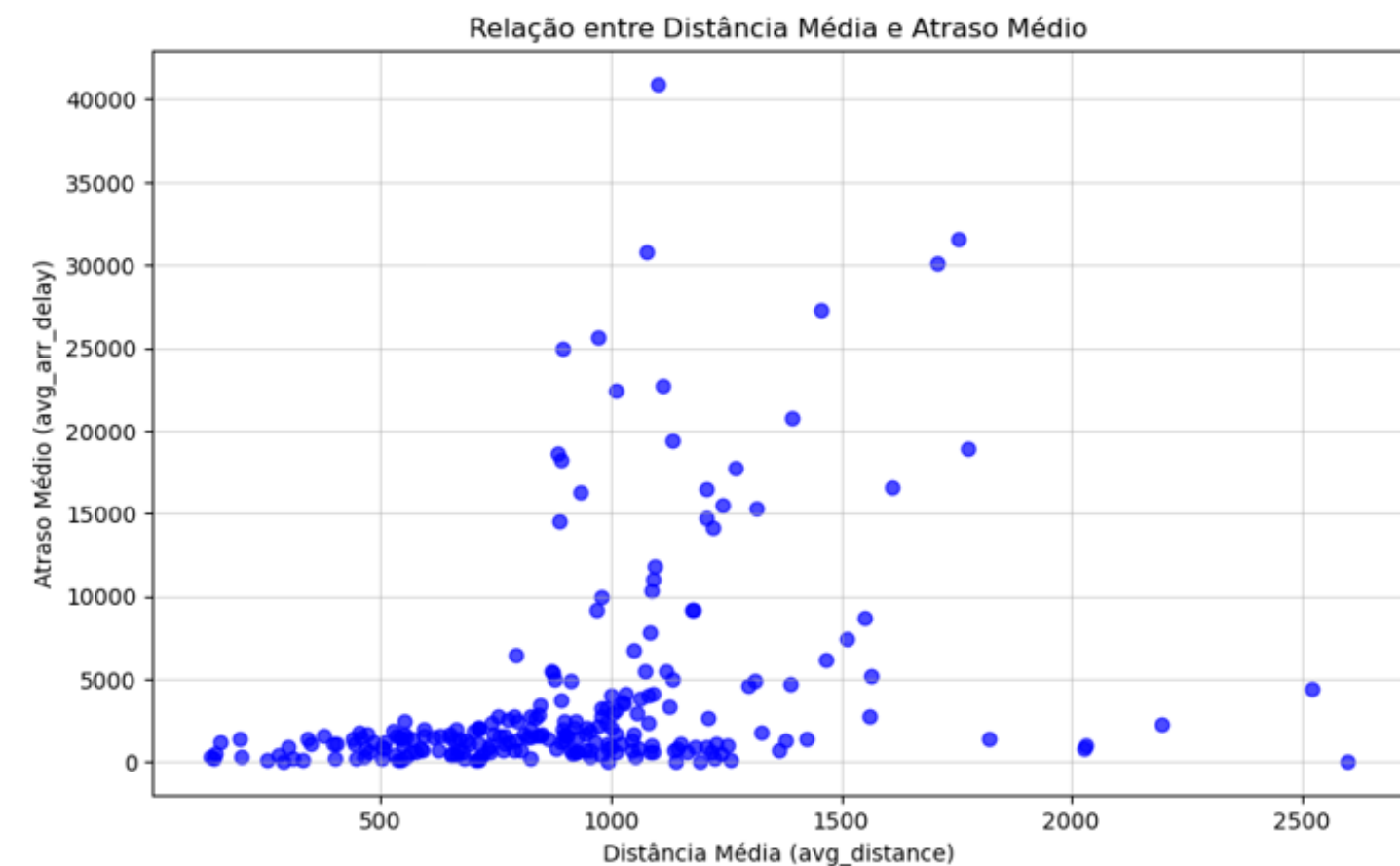
Dest	AveragePricePerTicket
DLG	229.4
OGS	89.05
CHS	194.78
FAY	221.2
ILM	254.68
ASE	263.0



Relação entre Distância Média e Atraso Médio

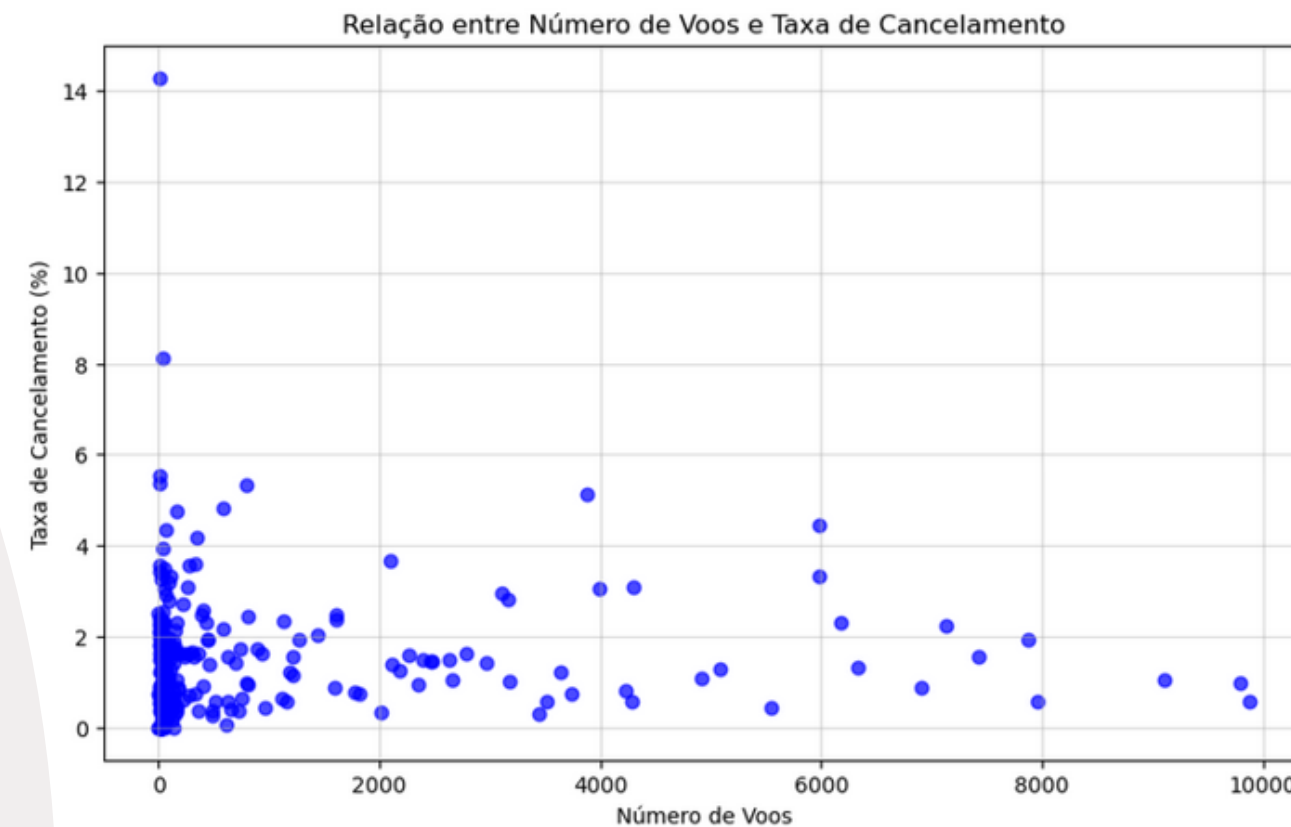
- A grande maioria dos voos apartir de certos aeroportos estão entre as 250 e a 1500 milhas
- O atraso médio da maioria dos voos é relativamente baixa, porém quando olhamos para voos acima das 750 milhas, começam a aparecer algumas exceções
- Os 5 aeroportos com voos mais compridos têm atrasos dentro da média

origin	avg_distance	avg_arr_delay
PPG	2599.0	26.0
HNL	2520.66	4457.86
OGG	2195.47	2238.64
KOA	2030.66	990.86
LIH	2027.98	842.29



Relação entre Número de Voos e Taxa de Cancelamento

- A maior parte dos aeroportos tem menos de 2000 voos e taxa de cancelamento abaixo de 4%
- Acima dos 2000 voos, os aeroportos em questão tem taxas de cancelamento bastante dispersas, um tem quase 0% e chegando a ultrapassar 5%
- CHS (Charleston international Airport) é um dos poucos com mais de 7000 voos e uma taxa de cancelamento acima dos 5%



origin	num_flights	cancel_rate
OGS	10	14.29
ILM	43	8.13
FAY	9	5.53
DLG	9	5.36
CHS	797	5.33

