Research and Model to understand the US aviation market

US FLIGHT ROUTES

DATA ANALYTICS

DAVID MAGRANER VILLALBA - DEC 2020

IRONHACK BERLIN

RESEARCH QUESTION

- There are more than 10 million flights within US per year, which makes it a massive industry.
- 6.5M jobs are directly supported by the air transport sector.
- Accounts for around 5% of the GDP of the US.
- Many companies offering similar services in national routes.

Are there any business opportunities for an Airline company by opening a brand new route within two unconnected US cities? If so, can we predict how many flights per month should we run the route with?

Continue in Tableau

CONSIDERATIONS & HYPOTHESES

- Restricted information availability.
  - We have routes, amount of flights per route, per month, per airline.
  - Merge with economic and demographic data to find out if we can explain the amount of flights between a pair of cities given the information we have.

ΔGDPShare -> ΔFlights

ΔGDPpc -> ΔFlights

ΔPopulation -> ΔFlights

Touristic areas -> ΔFlights

Summer months-> ΔFlights

Positive synergy between tourism and summer months.

MODEL SETUP

- Dummy variables:
  - Hub == 1 if the airport is a hub.
  - Away == 1 if in Alaska or Hawaii.
  - Within == 1 within the same state.
  - Summer == 1 in June, July or August.
  - Coast == 1 if the state has coast.
    - SumCoa == Summer * Coast, to control for the potential synergy.
- State GDP per capita and % of the US GDP Share.
- City population and distance between cities.
- Airport clusters (e.g. [LAX, SAN, SNA, ONT], [SFO, SMF, SJC, OAK], [JFK, EWR, LGA]...)

MODEL RESULTS + PREDICTION

OLS Regression Results

Dep. Variable:	Flights	R-squared (uncentered):	0.729
Model:	OLS	Adj. R-squared (uncentered):	0.729
Method:	Least Squares	F-statistic:	1.515e+04
Date:	Tue, 15 Dec 2020	Prob (F-statistic):	0.00
Time:	18:11:11	Log-Likelihood:	-3.9964e+05
No. Observations:	67613	AIC:	7.993e+05
Df Residuals:	67601	BIC:	7.994e+05
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Carrier	63.4309	0.298	212.526	0.000	62.846	64.016
Dist	-38.4432	0.662	-58.028	0.000	-39.742	-37.145
Hub	37.0673	0.748	49.538	0.000	35.601	38.534
Away	44.4714	2.178	20.419	0.000	40.203	48.740
Within	12.3053	1.612	7.633	0.000	9.146	15.465
Coast	21.9261	1.054	20.812	0.000	19.861	23.991
Summer	6.8082	1.193	5.707	0.000	4.470	9.146
GDP	6.5405	1.129	5.793	0.000	4.327	8.754
Pop	12.7450	0.182	70.163	0.000	12.389	13.101
Share	-4.5908	0.248	-18.517	0.000	-5.077	-4.105
Share2	0.2125	0.008	25.180	0.000	0.196	0.229
SumCoa	-3.0352	1.578	-1.923	0.054	-6.128	0.058

Omnibus:	28121.045	Durbin-Watson:	0.257
Prob(Omnibus):	0.000	Jarque-Bera (JB):	229735.339
Skew:	1.799	Prob(JB):	0.00
Kurtosis:	11.282	Cond. No.	1.23e+03

RESEARCH QUESTION

Are there any business opportunities for an Airline company to open a brand new route within two unconnected US cities? If so, can we predict how many flights per month should we run the route with?

ΔGDPShare -> ΔFlights

ΔGDPpc -> ΔFlights

ΔPopulation -> ΔFlights

Touristic areas -> ΔFlights

Summer months-> ΔFlights

Positive synergy between tourism and summer months.

MODEL RESULTS + PREDICTION

- We have taken 3 of the routes which are not connected yet.

	Route	Carrier	Dist	Hub	Away	Within	Coast	Summer	GDP	Pop	Share	Share2	SumCoa
0	MIA-PDX	1	2.70	1	0	0	1	0	1.047	1.51	6.27	39.31	0
1	MIA-PDX	1	2.70	1	0	0	1	1	1.047	1.51	6.27	39.31	1
2	BOS-SAT	1	1.76	1	0	0	1	0	1.442	1.05	11.43	130.64	0
3	BOS-SAT	1	1.76	1	0	0	1	1	1.442	1.05	11.43	130.64	1
4	PIT-MCI	1	0.77	0	0	0	0	0	1.151	2.64	5.43	29.49	0
5	PIT-MCI	1	0.77	0	0	0	0	1	1.151	2.64	5.43	29.49	0

```
model.predict(xp)

0    24.289652
1    28.062687
2    52.866719
3    56.639753
4    56.343079
5    63.151320
```

FUTURE IMPROVEMENTS + LEARNING

- Improvement of the model:
  - More reliable information.
    - Passengers per route to interpret demand; we know #flights but what is the size?
    - Pricing will definitely influence the demand for flights. Not enough data.
  - We could create various models instead of a model with several dummies.
  - Carriers and Flights can be correlated both ways, we should control for that.
- Learning:
  - Tableau Dashboards, Python. Machine Learning and Feature Engineering.
  - Time management and focusing on establishing an idea and providing an MVP.