



AIRTICKETS – FINAL PROJECT

Tomancová Nikola



OBSAH



POPIS
DATASETU
A ZKOUMANÉ
OBLASTI



PŘÍPRAVA DAT
A ANALÝZA
V POWER BI



DATA MINING –
ANALÝZA V
CLEVERMINERU



POPIS ŘEŠENÉ OBLASTI

- Analýza letenek na vybrané trasy ([Expedia.com](https://www.expedia.com))
- Sledované období: **17. dubna až 5. května 2022**
- Sledovaná oblast: **USA (vnitrostátní lety)**
- Pozorování rozdílů např. pro různé letecké společnosti, cestovní třídy, dny v týdnu, destinace, meteorologické podmínky...



CÍLE ANALÝZY

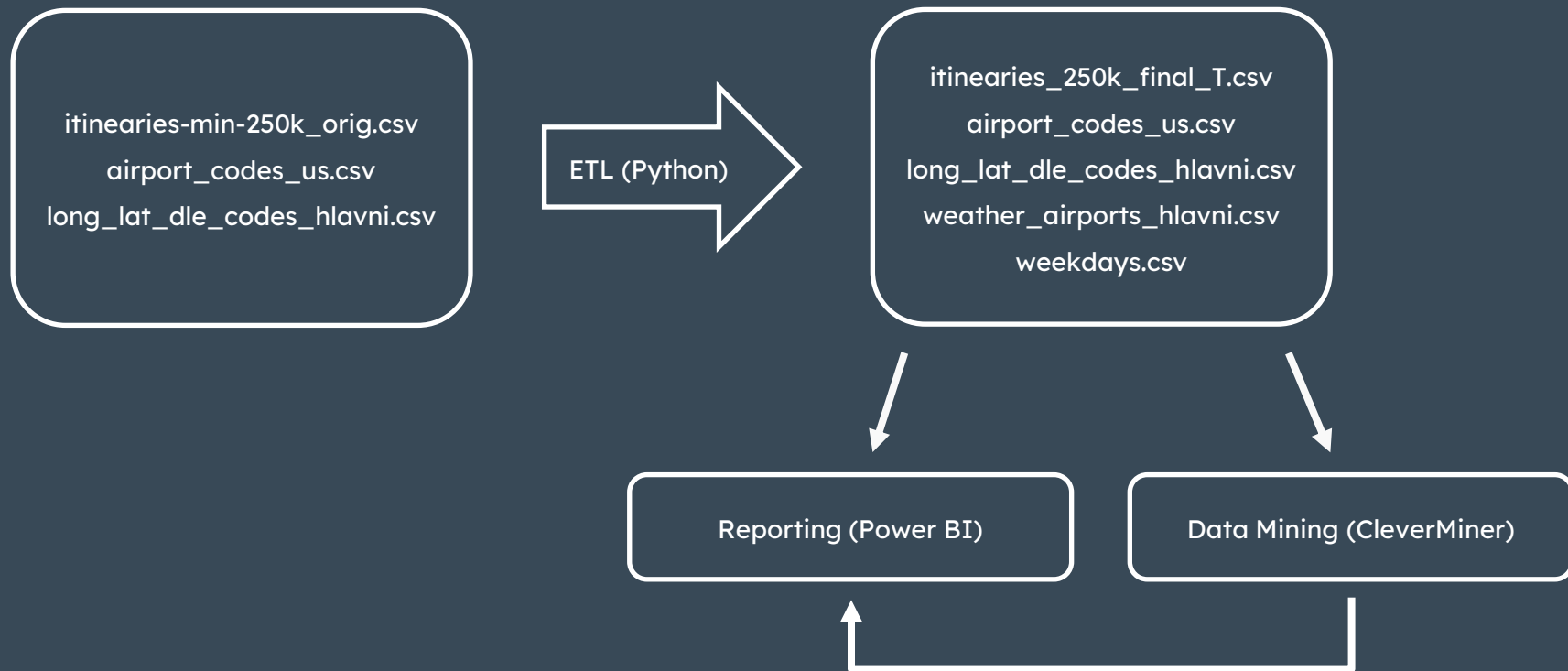
Účelem této práce je pomoci s průzkumem pro začínající leteckou společnost, která zkoumá, jakým způsobem oceňují své letenky konkurenční aerolinky a také například zvažuje, na jaké trasy by bylo vhodné se soustředit.

To tedy mimo jiné znamená:

- Identifikaci faktorů ovlivňujících cenu letenky
- Hledání zajímavých souvislostí v datech
- Návrh využití poznatků v praxi



ARCHITEKTURA ŘEŠENÍ



PŘÍPRAVA & VIZUALIZACE DAT





AIR TICKETS TABLE



```
df = pd.read_csv('itineraries-min-250k_orig.csv',
                  sep=",", low_memory= False )
df.columns

Index(['legId', 'searchDate', 'flightDate', 'startingAirport',
       'destinationAirport', 'fareBasisCode', 'travelDuration', 'elapsedDays',
       'isBasicEconomy', 'isRefundable', 'isNonStop', 'baseFare', 'totalFare',
       'seatsRemaining', 'totalTravelDistance',
       'segmentsDepartureTimeEpochSeconds', 'segmentsDepartureTimeRaw',
       'segmentsArrivalTimeEpochSeconds', 'segmentsArrivalTimeRaw',
       'segmentsArrivalAirportCode', 'segmentsDepartureAirportCode',
       'segmentsAirlineName', 'segmentsAirlineCode',
       'segmentsEquipmentDescription', 'segmentsDurationInSeconds',
       'segmentsDistance', 'segmentsCabinCode'],
      dtype='object')
```

df.columns

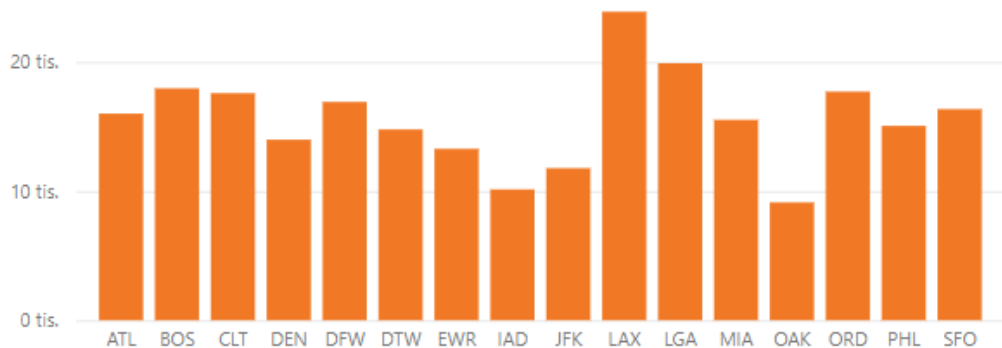
```
Index(['searchDate', 'flightDate', 'startingAirport', 'destinationAirport',
       'elapsedDays', 'isBasicEconomy', 'isNonStop', 'baseFare', 'totalFare',
       'seatsRemaining', 'totalTravelDistance', '1st_segmentDurationInSeconds',
       '2nd_segmentDurationInSeconds', '3rd_segmentDurationInSeconds',
       '4th_segmentDurationInSeconds', 'CabinCodeSummary',
       'AirlineNameSummary', 'AirlineNameCount', 'EquipmentDescriptionSummary',
       'NumberOfTransfers', 'DaysToFlight', 'TotalDurationInSeconds',
       'TotalTravelDuration', 'WaitingTime', 'increaseFare', 'dep_time_cat',
       'arr_time_cat', 'DaysToFlight_cat', 'Tot_Trav_Duration_cat',
       'Wait_time_cat', 'totalTravelDistance_cat', 'totalFare_cat',
       'increaseFare_cat', 'arrivalDate', 'a_dep_index', 'a_arr_index',
       'TotalTravelDurationInHours', 'WaitingTimeInHours'],
      dtype='object')
```

```
df.to_csv(path_or_buf='itineraries_250k_final3.csv')
```

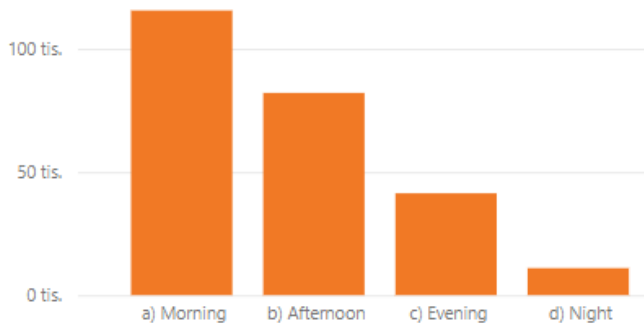
Airport_Code Airport_City

ATL	Atlanta
BOS	Boston
CLT	Charlotte
DEN	Denver
DFW	Dallas
DTW	Detroit
EWR	Newark
IAD	Dulles, DC
JFK	New York
LAX	Los Angeles
LGA	New York
MIA	Miami
OAK	Oakland
ORD	Chicago
PHL	Philadelphia
SFO	San Francisco

Rámcový profil atributu startingAirport



Rámcový profil atributu dep_time_cat



00:00 – 05:59

→ d) Night

06:00 – 11:59

→ a) Morning

12:00 – 17:59

→ b) Afternoon

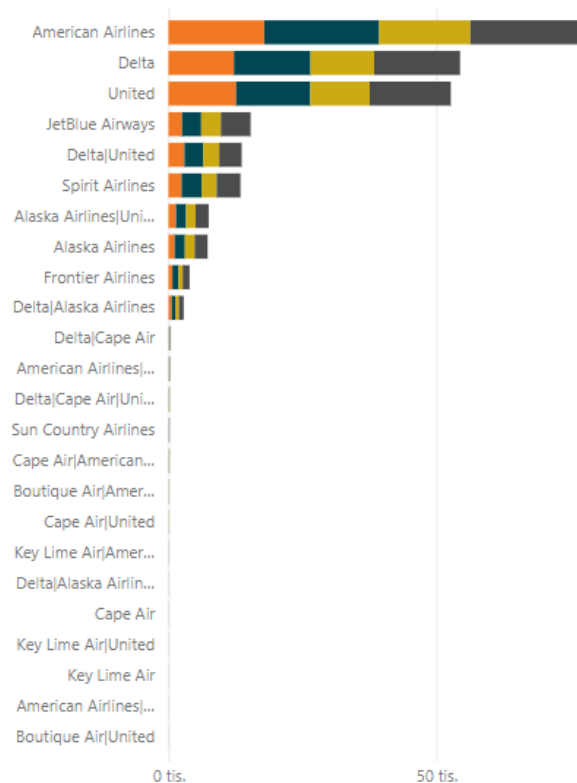
18:00 – 23:59

→ c) Evening



Rámcový profil atributu AirlineNameSummary

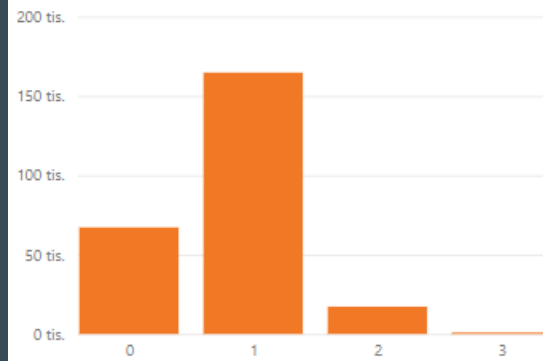
DaysToFlight_cat a) < 4 Days b) 4 - 7 Days c) 8 - 10 Days d) 11+ Days



Rámcový profil atributu Tot_Trav_Duration_cat



Rámcový profil atributu NumberOfTransfers

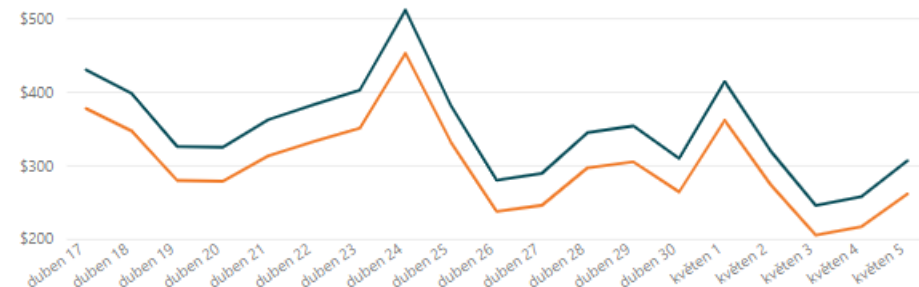


CabinCodeSummary

business	business coach	business first	coach	coach first	coach premium coach	first	premium coach
----------	----------------	----------------	-------	-------------	---------------------	-------	---------------

Průměr z: baseFare a Průměr z: totalFare podle kategorie Měsíc a Den

● Průměr z: baseFare ● Průměr z: totalFare



\$302

Průměr z: baseFare

\$350

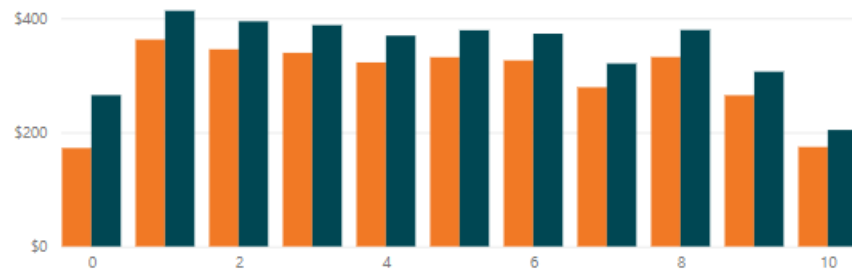
Průměr z: totalFare

1,16

totalXbase

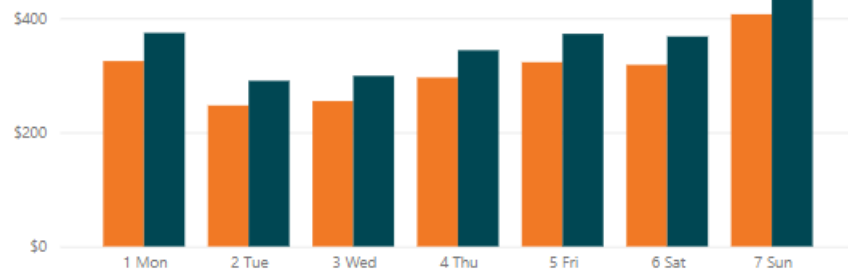
Průměr z: baseFare a Průměr z: totalFare podle kategorie seatsRemaining

● Průměr z: baseFare ● Průměr z: totalFare



Průměr z: baseFare a Průměr z: totalFare podle kategorie Weekday

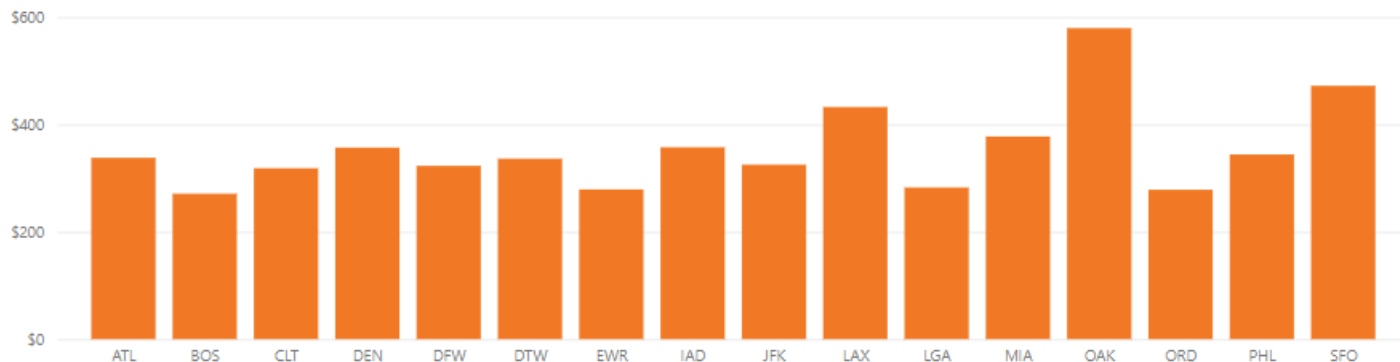
● Průměr z: baseFare ● Průměr z: totalFare



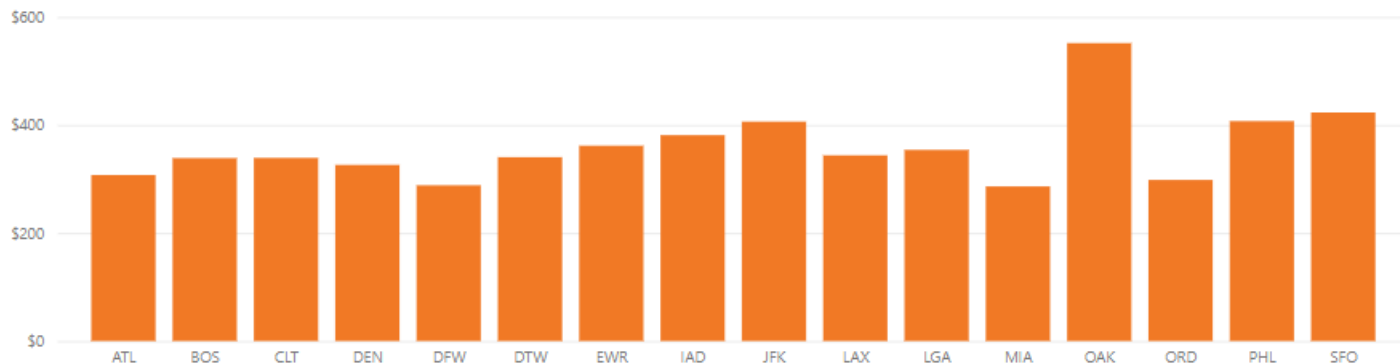
Weekday

1 Mon	2 Tue	3 Wed	4 Thu	5 Fri	6 Sat	7 Sun
-------	-------	-------	-------	-------	-------	-------

Průměr z: totalFare podle kategorie startingAirport



Průměr z: totalFare podle kategorie destinationAirport

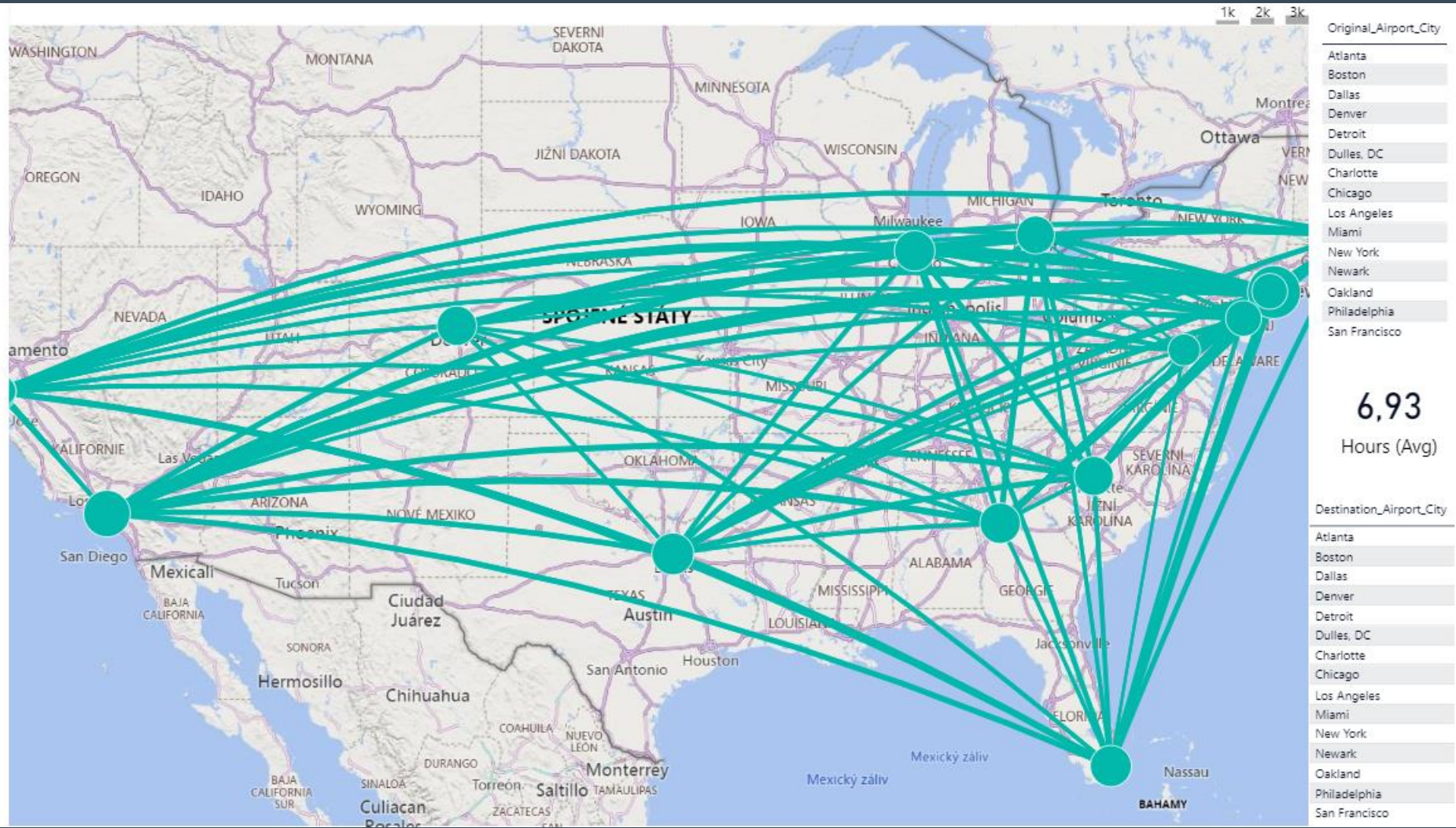


Airport_Code Airport_City

ATL	Atlanta
BOS	Boston
CLT	Charlotte
DEN	Denver
DFW	Dallas
DTW	Detroit
EWB	Newark
IAD	Dulles, DC
JFK	New York
LAX	Los Angeles
LGA	New York
MIA	Miami
OAK	Oakland
ORD	Chicago
PHL	Philadelphia
SFO	San Francisco

\$350

Průměr z: totalFare



ZÁVĚR BI ANALÝZY

- V rámci BI analýzy se podařilo odhalit atributy významně ovlivňující cenu letenky.
 - Den v týdnu
 - Počáteční/cílová destinace
 - Délka cesty
 - Počet dní do odletu
 - Denní doba odletu
 - Počet zbývajících míst





DATA SCIENCE



Detailnější zkoumání souvislostí mezi atributy na základě výsledků předchozí analýzy v Power BI



Příprava dat pro Cleverminer

- Potřebné atributy
byly sloučeny do
jedné datové matice

```
weather = pd.read_csv('weather_airports_hlavni.csv')
weather.drop(labels=['Unnamed: 0', 'time', 'tavg', 'tmin', 'tmax', 'prcp', 'snow',
                    'wdir', 'wspd', 'wpgt', 'pres', 'tsun', 'Airport_Code'],
            axis = 'columns', inplace = True)
weekdays = pd.read_csv('weekdays.csv')
df = pd.read_csv('itineraries_250k_final3.csv', low_memory= False )
df = df[['a_dep_index', 'a_arr_index', 'flightDate',
        'startingAirport', 'destinationAirport', 'elapsedDays',
        'isBasicEconomy', 'isNonStop',
        'seatsRemaining',
        'CabinCodeSummary', 'NumberOfTransfers',
        'dep_time_cat', 'arr_time_cat', 'Tot_Trav_Duration_cat',
        'Wait_time_cat', 'increaseFare_cat',
        'totalFare_cat', 'totalTravelDistance_cat', 'AirlineNameSummary',
        'AirlineNameCount', 'EquipmentDescriptionSummary', 'DaysToFlight_cat']]
matice = pd.merge(df, weather, how="left", left_on='a_arr_index', right_on = 'w_index')
matice = pd.merge(matice, weekdays, how="left", left_on='flightDate', right_on = 'Date')
matice.drop(labels=['a_dep_index', 'w_index', 'Date'], axis = 'columns', inplace = True)
```

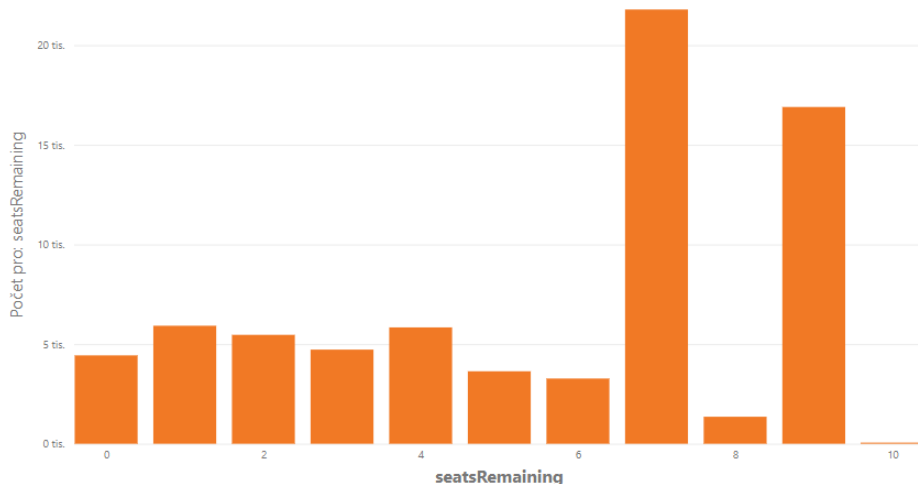


I) POSÍLENÍ VYTÍŽENÝCH SPOJŮ

- *Bylo by vhodné na některých trasách přidat lety?*
- Jinými slovy, existuje nějaká trasa, na které se spolehlivostí alespoň 40 % budou 11 dní před odletem zbývat maximálně 2 poslední volná sedadla splňující zákazníkem zadané parametry?

Seats Remaining - 11 Days to Flight

Počet pro: seatsRemaining podle kategorie seatsRemaining



4FT-MINER

CleverMiner task processing summary:

Task type : 4ftMiner
Number of verifications : 507
Number of rules : 1
Total time needed : 00h 00m 02s
Time of data preparation : 00h 00m 02s
Time of rule mining : 00h 00m 00s

List of rules:

RULEID	BASE	CONF	AAD	Rule
1	104	0.446	+1.531	DaysToFlight_cat(d) 11+ Days) & startingAirport(DTW) & destinationAirport(EWR) => seatsRemaining(0 1) ---

Rule id : 1

Base : 104 Relative base : 0.000 CONF : 0.446 AAD : +1.531 BAD : -1.531

Cedents:

antecedent : DaysToFlight_cat(d) 11+ Days) & startingAirport(DTW) & destinationAirport(EWR)
succedent : seatsRemaining(0 1)
condition : ---

Fourfold table

	S	¬S
A	104	129
¬A	43982	205785

```
clm = cleverminer(df=matice,proc='4ftMiner',
  quantifiers= {'conf':0.4, 'Base':50},
  ante ={
    'attributes':[
      {'name': 'DaysToFlight_cat', 'type':'one', 'value': 'd) 11+ Days'},
      {'name': 'startingAirport', 'type': 'subset', 'minlen': 1, 'maxlen':1},
      {'name': 'destinationAirport', 'type': 'subset', 'minlen': 1, 'maxlen': 1}
    ], 'minlen':3, 'maxlen':3, 'type':'con'},
  succ ={
    'attributes':[
      {'name': 'seatsRemaining', 'type':'lcut', 'minlen': 1, 'maxlen':2},
      {'name': 'seatsRemaining', 'type': 'subset', 'minlen': 1, 'maxlen':1},
    ], 'minlen':1, 'maxlen':1, 'type':'con'},
  )

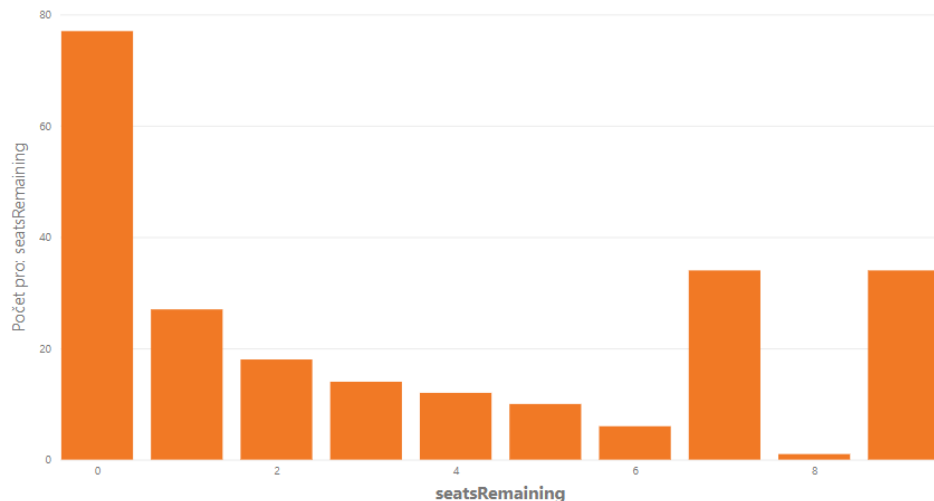
clm.print_summary()
clm.print_rulelist()
clm.print_rule(1)
```

- $\text{Konfidence} = 104 / (104 + 129) = 0,446$
- 44,6 % letenek splňujících DaysToFlight_cat(d) 11+ Days) & startingAirport(DTW) & destinationAirport(EWR) splňuje i seatsRemaining(0 1)

POSÍLENÍ VYTÍŽENÝCH SPOJŮ – INTERPRETACE

Seats Remaining - 11 Days to Flight

Počet pro: seatsRemaining podle kategorie seatsRemaining



startingAirport

ATL	DFW	MIA	SFO
BOS	DTW	OAK	
CLT	IAD	ORD	
DEN	LAX	PHL	

destinationAirport

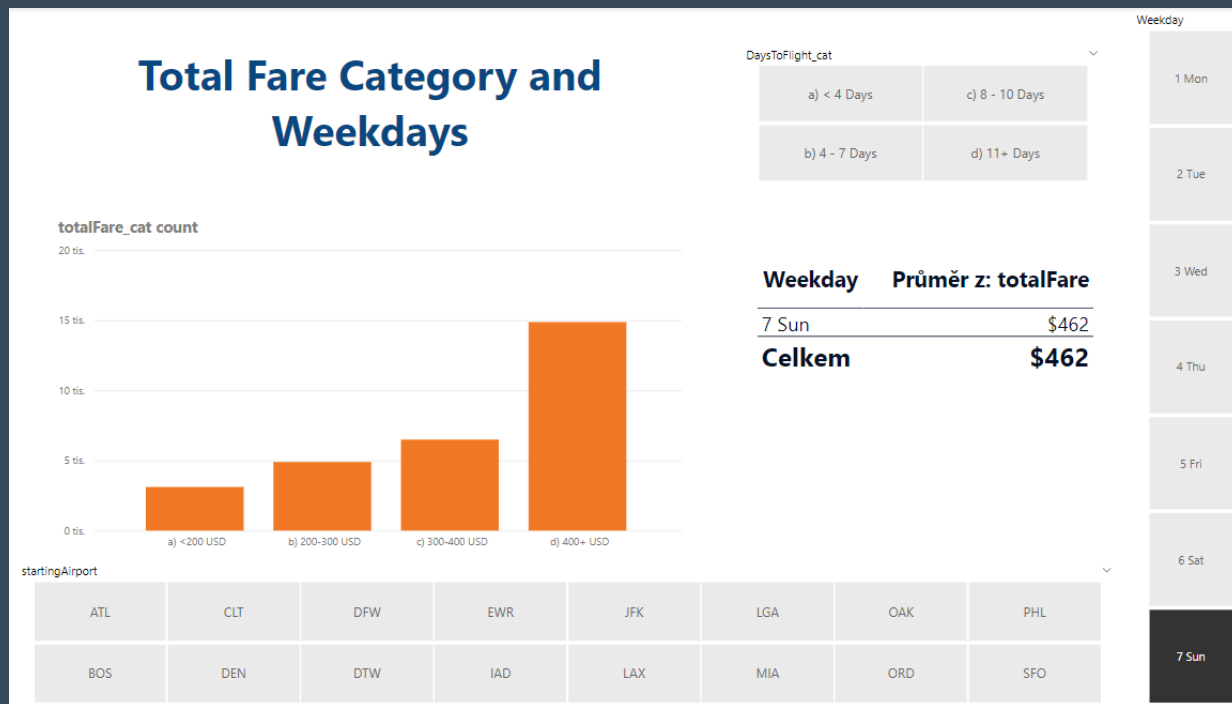
ATL	DFW	LAX	ORD
BOS	EWR	LGA	PHL
CLT	IAD	MIA	SFO
DEN	JFK	OAK	

- Získané pravidlo říká, že téměř 45 % spojů z Detroitu do Newarku bude mít už 11 dní předem téměř vyprodáno (hodnota SeatRemaining 0 nebo 1)
- Bylo by vhodné zvážit posílení této trasy, zjevně je po ní vysoká poptávka.



IIa) CENY LETENEK & DNY V TÝDNU (NEDĚLE)

- Existuje nějaká kombinace trasy, meteorologických údajů, denní doby odletu, počtu zbývajících volných míst v letadle a počtu dní do odletu tak, že pro tuto kombinaci bude histogram cen nedělních letenek klesající?



CF-MINER

CleverMiner task processing summary:

Task type : CFMiner

Number of verifications : 19665

Number of rules : 1839

Total time needed : 00h 00m 15s

Time of data preparation : 00h 00m 02s

Time of rule mining : 00h 00m 12s

List of rules:

RULEID	BASE	S_UP	S_DOWN	Condition
--------	------	------	--------	-----------

1	712	0	3	Weekday(7 Sun) & DaysToFlight_cat(a) < 4 Days) & startingAirport(ORD)
2	654	0	3	Weekday(7 Sun) & DaysToFlight_cat(a) < 4 Days) & startingAirport(ORD) & prcp_cat(a) 0-1)
3	592	0	3	Weekday(7 Sun) & DaysToFlight_cat(d) 11+ Days) & startingAirport(BOS)
4	801	0	3	Weekday(7 Sun) & startingAirport(LGA) & tavg_cat(d) 20-30)

Rule id : 1

Base : 712 Relative base : 0.003 Steps UP (consecutive) : 0 Steps DOWN (consecutive) : 3

6 Histogram relative maximum : 0.367 Histogram relative minimum : 0.149

Condition : Weekday(7 Sun) & DaysToFlight_cat(a) < 4 Days) & startingAirport(ORD)

Categories in target variable ['a) >200 USD', 'b) 200-300 USD', 'c) 300-400 USD', 'd) 400+ USD']

Histogram [261, 219, 126, 106]

```
clm = cleverminer(df=matice,target='totalFare_cat',proc='CFMiner',
                 quantifiers= {'S_Down':3, 'Base':500},
                 cond ={
                     'attributes':[
                         {'name': 'Weekday', 'type': 'one', 'value': '7 Sun'},
                         {'name': 'dep_time_cat', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
                         {'name': 'seatsRemaining', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
                         {'name': 'DaysToFlight_cat', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
                         {'name': 'startingAirport', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
                         {'name': 'destinationAirport', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
                         {'name': 'tavg_cat', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
                         {'name': 'prcp_cat', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
                         {'name': 'wspd_cat', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
                     ], 'minlen':2, 'maxlen':9, 'type':'con'}
                 )

clm.print_summary()
clm.print_rulelist()
clm.print_rule(1)
```

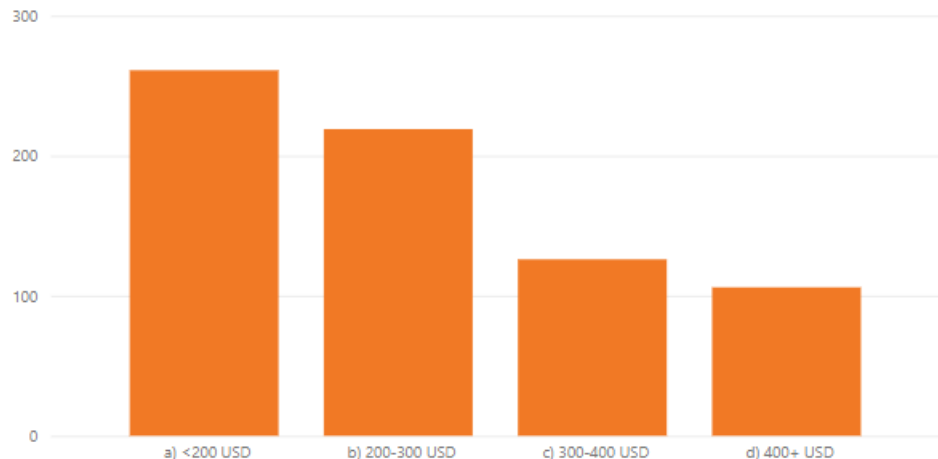
CENY LETENEK & NEDĚLE – INTERPRETACE

- Celkem 4 histogramy nedělních odletů splňující zadané parametry (ale druhý z histogramů již vyplývá z prvního).
- Například nedělní letenky z Chicaga bývají neobvykle levné, zbývají-li nejvýše 3 dny do odletu.
- Naopak nedělní letenky z Bostonu lze sehnat nezvykle levně, nakupujeme-li s dostatečným předstihem.



Total Fare Category and Weekdays

totalFare_cat count



startingAirport

ATL	CLT	DFW	EWK	JFK	LGA	OAK	PHL
BOS	DEN	DTW	IAD	LAX	MIA	ORD	SFO

DaysToFlight_cat

a) < 4 Days	c) 8 - 10 Days
b) 4 - 7 Days	d) 11+ Days

Weekday Průměr z: totalFare

7 Sun \$286

Celkem \$286

Weekday

1 Mon

2 Tue

3 Wed

4 Thu

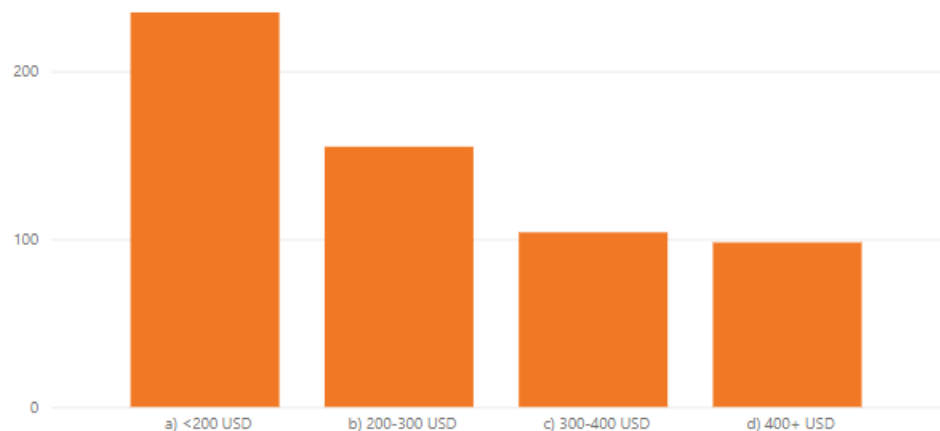
5 Fri

6 Sat

7 Sun

Total Fare Category and Weekdays

totalFare_cat count



startingAirport

ATL	CLT	DFW	EWL	JFK	LGA	OAK	PHL
BOS	DEN	DTW	IAD	LAX	MIA	ORD	SFO

DaysToFlight_cat

a) < 4 Days	c) 8 - 10 Days
b) 4 - 7 Days	d) 11+ Days

Weekday Průměr z: totalFare

7 Sun	\$267
-------	-------

Celkem	\$267
---------------	--------------

Weekday

1 Mon

2 Tue

3 Wed

4 Thu

5 Fri

6 Sat

7 Sun

III) CENY LETENEK & AEROLINKY

- BI analýza za obecně levnější aerolinku označila například American Airlines, za velmi drahou v porovnání s ostatními zase jednoznačně Alaska Airlines. V rámci mapování konkurence je proto nyní cílem najít takové trasy, na kterých lze létat s některými aerolinkami levněji než s American Airlines, resp. draž než s Alaska Airlines.

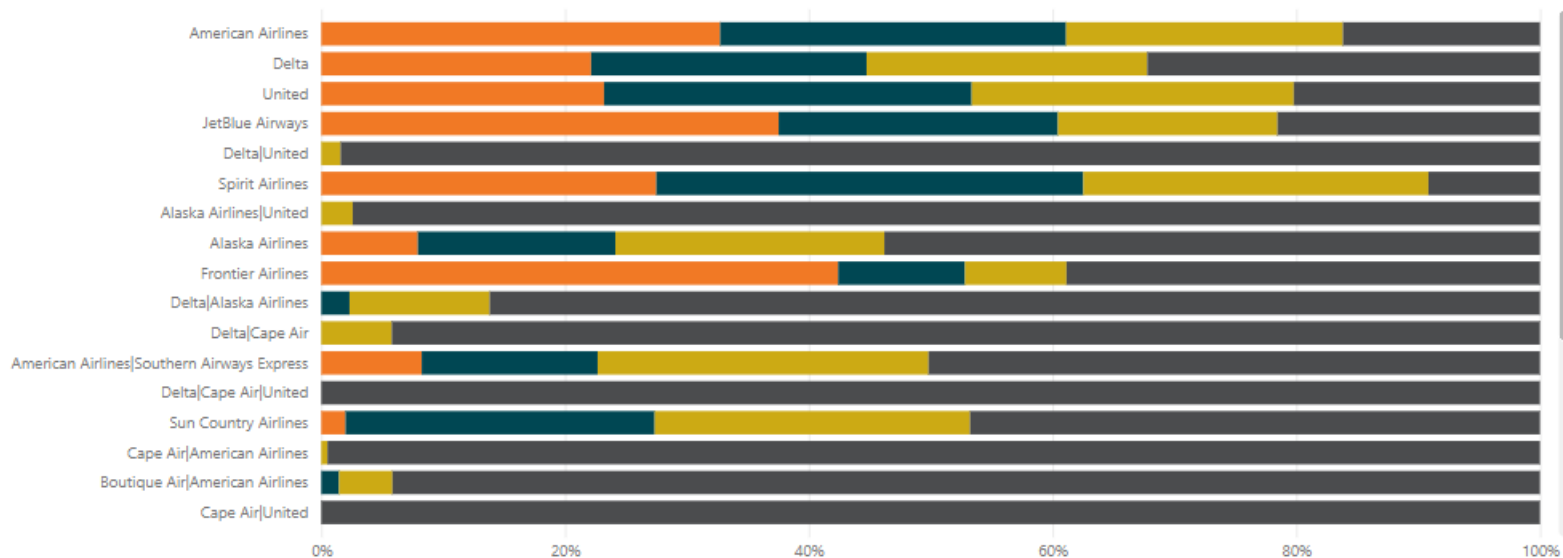


IIIa) CENY LETENEK & AEROLINKY

- *Je pro některé trasy relativní četnost levných letenek (<200 USD) více než 1.5x menší pro American Airlines oproti některé z jiných aerolinek (resp. jiné skupiny aerolinek)?*

AirlineNameSummary & TotalFare_cat (Relative Values)

totalFare_cat ● a) <200 USD ● b) 200-300 USD ● c) 300-400 USD ● d) 400+ USD



SD4FT-MINER

Number of verifications : 4959
Number of rules : 18
Total time needed : 00h 00m 08s
Time of data preparation : 00h 00m 02s
Time of rule mining : 00h 00m 05s

List of rules:

RULEID	BASE1	BASE2	RatioConf	DeltaConf	Rule
1	133	50	2.010	+0.292	destinationAirport(LAX) & startingAirport(SFO) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Alaska Airlines) x AirlineNameSummary(American Airlines)
2	190	103	1.585	+0.206	destinationAirport(ATL) & startingAirport(DFW) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Delta) x AirlineNameSummary(American Airlines)
3	63	62	1.746	+0.131	destinationAirport(DEN) & startingAirport(ATL) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Delta) x AirlineNameSummary(American Airlines)
4	106	50	3.332	+0.248	destinationAirport(DFW) & startingAirport(ORD) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Delta) x AirlineNameSummary(American Airlines)
5	64	61	1.742	+0.257	destinationAirport(IAD) & startingAirport(BOS) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Delta) x AirlineNameSummary(American Airlines)
6	120	50	2.129	+0.326	destinationAirport(LAX) & startingAirport(SFO) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Delta) x AirlineNameSummary(American Airlines)
7	87	134	2.086	+0.401	destinationAirport(ATL) & startingAirport(PHL) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Frontier Airlines) x AirlineNameSummary(American Airlines)
8	52	62	1.865	+0.185	destinationAirport(EWR) & startingAirport(MIA) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(JetBlue Airways) x AirlineNameSummary(American Airlines)
9	111	101	1.632	+0.094	destinationAirport(LAX) & startingAirport(JFK) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(JetBlue Airways) x AirlineNameSummary(American Airlines)
10	98	134	1.782	+0.289	destinationAirport(ATL) & startingAirport(PHL) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Spirit Airlines) x AirlineNameSummary(American Airlines)
11	93	113	1.590	+0.233	destinationAirport(DFW) & startingAirport(ATL) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Spirit Airlines) x AirlineNameSummary(American Airlines)
12	66	121	1.838	+0.157	destinationAirport(DFW) & startingAirport(LAX) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Spirit Airlines) x AirlineNameSummary(American Airlines)
13	58	50	3.188	+0.233	destinationAirport(DFW) & startingAirport(ORD) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Spirit Airlines) x AirlineNameSummary(American Airlines)
14	83	98	1.748	+0.254	destinationAirport(DTW) & startingAirport(ATL) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(Spirit Airlines) x AirlineNameSummary(American Airlines)
15	173	137	1.938	+0.367	destinationAirport(DFW) & startingAirport(EWR) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(United) x AirlineNameSummary(American Airlines)
16	73	50	2.932	+0.206	destinationAirport(DFW) & startingAirport(ORD) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(United) x AirlineNameSummary(American Airlines)
17	141	61	1.858	+0.297	destinationAirport(IAD) & startingAirport(BOS) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(United) x AirlineNameSummary(American Airlines)
18	230	50	2.689	+0.488	destinationAirport(LAX) & startingAirport(SFO) => totalFare_cat(a) >200 USD --- : AirlineNameSummary(United) x AirlineNameSummary(American Airlines)

Rule id : 1

Base1 : 133 Base2 : 50 Relative base 1 : 0.001 Relative base 2 : 0.000 CONF1 : 0.581 CONF2 : +0.289 Delta Conf : +0.292 Ratio Conf : +2.010

Cedents:

antecedent : destinationAirport(LAX) & startingAirport(SFO)
succedent : totalFare_cat(a) >200 USD
condition : ---
first set : AirlineNameSummary(Alaska Airlines)
second set : AirlineNameSummary(American Airlines)

Fourfold tables:

FRST	S	~S	SCND	S	~S
---	----	----	---	----	----
A	133	96	A	50	123
---	----	----	---	----	----
~A	446	6632	~A	25084	51540

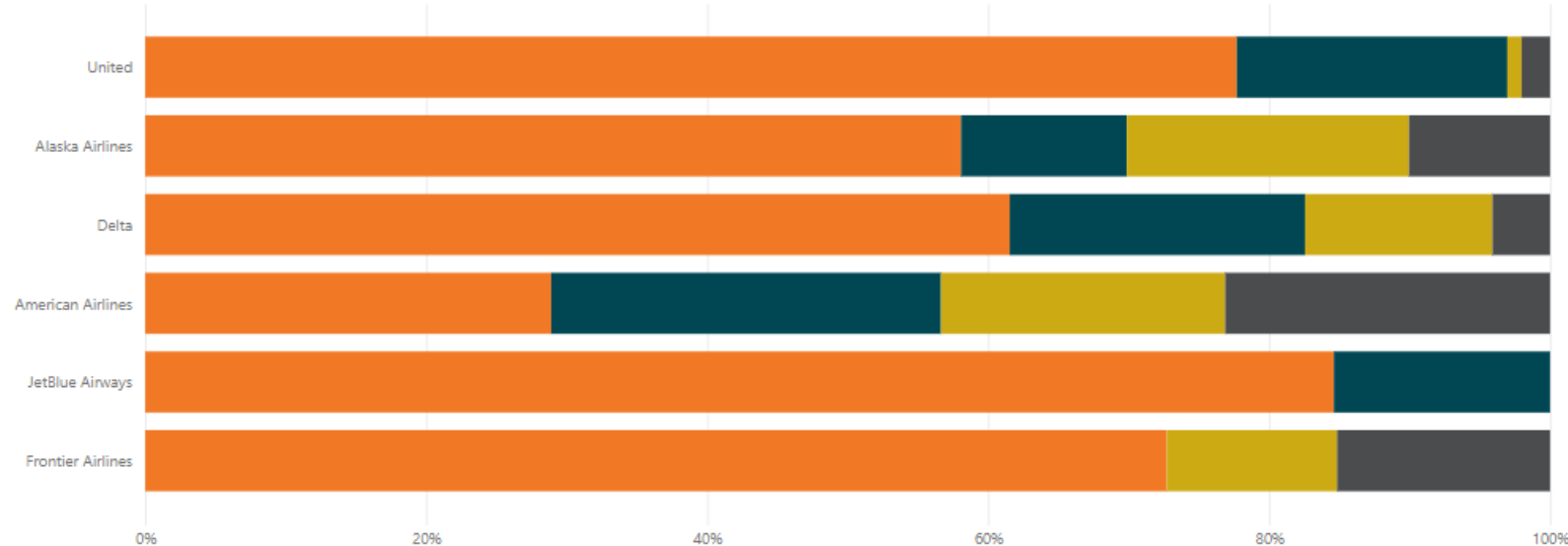
```
clm = cleverminer(df=matice,proc='SD4ftMiner',
quantifiers= {'Base1':50, 'Base2':50, 'Ratioconf' : 1.5},
ante ={'attributes':[
{'name': 'destinationAirport', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
{'name': 'startingAirport', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
], 'minlen':2, 'maxlen':2, 'type':'con'},
succ ={'attributes':[
{'name': 'totalFare_cat', 'type': 'one', 'value':'a' <200 USD}
], 'minlen':1, 'maxlen':1, 'type':'con'},
frst ={'attributes':[
{'name': 'AirlineNameSummary', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
], 'minlen':1, 'maxlen':1, 'type':'con'},
scnd ={'attributes':[
{'name': 'AirlineNameSummary', 'type': 'one', 'value': 'American Airlines'},
], 'minlen':1, 'maxlen':1, 'type':'con'},
)
```

- $\text{RatioConf} = \frac{133/(133+96)}{50/(50+123)} = \frac{0,581}{0,289} = 2,010$
- Relativní četnost letenek ze SFO do LAX, které stojí méně než 200 USD, je u Alaska Airlines 2,01x vyšší než u American Airlines.

Airlines and Flight Prices

AirlineNameSummary & TotalFare_cat (Relative Values)

totalFare_cat a) <200 USD b) 200-300 USD c) 300-400 USD d) 400+ USD



startingAirport

ATL	CLT	DFW	EWR	JFK	MIA	ORD	SFO
BOS	DEN	DTW	IAD	LGA	OAK	PHL	

destinationAirport

ATL
BOS
CLT
DEN
DFW
DTW
EWR
IAD
JFK
LAX
LGA
MIA
OAK
ORD
PHL

IV) PREDIKTIVNÍ MODELY – RANDOM FOREST

- Rozdělení dat na trénovací a testovací v poměru 8:2
- Parametry RandomForestClassifieu byly nastaveny následovně:
 - n_estimators jako počet stromů v lese roven 50
 - Random_state pro zajištění reproducibility roven 42
 - Ostatní parametry ponechány na defaultních hodnotách

```
rf_classifier = RandomForestClassifier(n_estimators=50, random_state=42)

rf_classifier.fit(X_train, y_train)

y_pred_rf = rf_classifier.predict(X_test)

accuracy_rf = accuracy_score(y_test, y_pred_rf)
print(f"Accuracy: {accuracy_rf}")

Accuracy: 0.85394
```

```
report_rf = classification_report(y_test, y_pred_rf)
print(report_rf)
```

	precision	recall	f1-score	support
a) >200 USD	0.89	0.90	0.90	12233
b) 200-300 USD	0.81	0.81	0.81	12015
c) 300-400 USD	0.79	0.77	0.78	10689
d) 400+ USD	0.90	0.91	0.91	15063
accuracy			0.85	50000
macro avg	0.85	0.85	0.85	50000
weighted avg	0.85	0.85	0.85	50000





VYHODNOCENÍ PREDIKTIVNÍCH MODELŮ

RANDOM FOREST

Accuracy: 85.39 %

AUC: 0.97

Nízká časová náročnost

RF TUNED

Accuracy: 85.46 %

AUC: 0.97

Vysoká časová náročnost





ZÁVĚR



Shrnutí nově nabytých znalostí z předchozích
analýz

ZÁVĚR

- Stanovení cen letenek:
 - podobně jako ostatní aerolinky
 - X opačně než ostatní aerolinky,
 - predikční model jako prvotní odhad.
- Objevení potenciálně přetížených tras.





DĚKUJI ZA
POZORNOST!

RESOURCES I

- <https://air.flyingway.com/books/xls/airport-codes.xls>
- <https://www.kaggle.com/datasets/justinmitchel/flightprices-min?resource=download&select=itineraries-min-250k.csv>
- https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FLL&QO_fu146_a_nzr=
- <https://dev.meteostat.net/python/>
- <https://www.expedia.com/Flights>
- https://www.budgetair.com/en_ca/blog/what-are-the-different-cabin-classes-on-airplanes
- <https://simpleflying.com/how-airline-ticket-pricing-works/>



RESOURCES II

- <https://www.flightapi.io/blog/airline-pricing-strategies/>
- <https://www.w3.org/TR/NOTE-datetime>
- <https://www.chmi.cz/files/portal/docs/meteo/om/sivs/dest.html>
- <https://www.rmets.org/metmatters/beaufort-wind-scale>

PHOTOS, ICONS AND TEMPLATE:

- <https://slidesgo.com/theme/plane-flying-in-the-sky#search-Airplane&position-10&results-70>
- [Icon Pack | Aviation \(flaticon.com\)](#)
- [Free Photo | Plane flying in sunset sky \(freepik.com\)](#)
- [Free Photo | Low angle tall chimney and airplane \(freepik.com\)](#)
- [Free Photo | Daytime skyscape \(freepik.com\)](#)



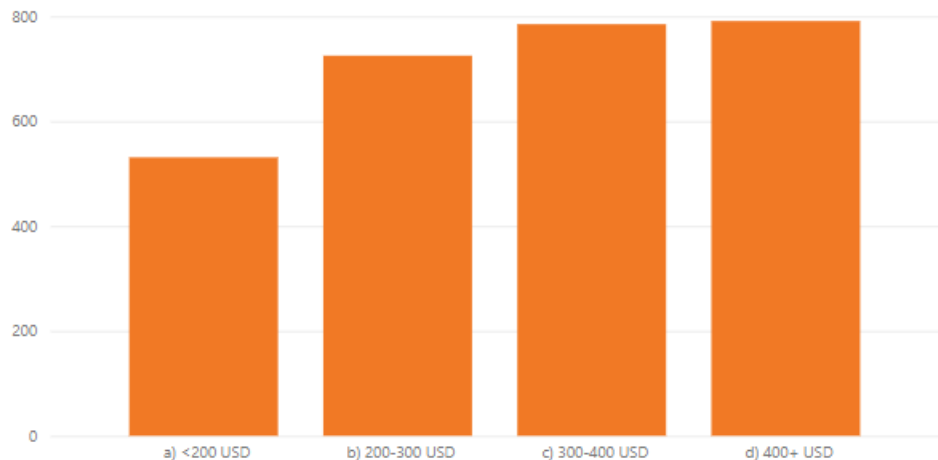
IIb) CENY LETENEK & ÚTERÝ – INTERPRETACE

- Úterní letenky od letecké společnosti Delta bývají dražší.
- Další zajímavé nalezené pravidlo říká, že v úterý bývají drahé také večerní letenky kupované maximálně 3 dny před odletem.



Total Fare Category and Weekdays

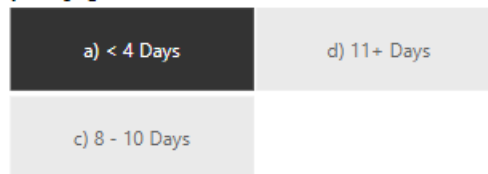
totalFare_cat count



startingAirport

ATL	CLT	DFW	EWR	JFK	LGA	OAK	PHL
BOS	DEN	DTW	IAD	LAX	MIA	ORD	SFO

DaysToFlight_cat



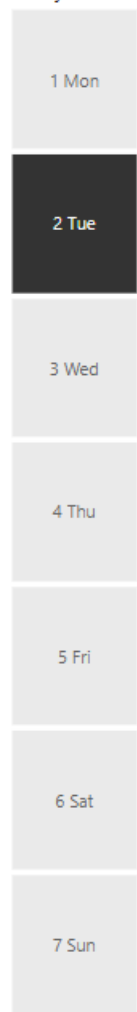
Weekday Průměr z: totalFare

2 Tue	\$362
Celkem	\$362

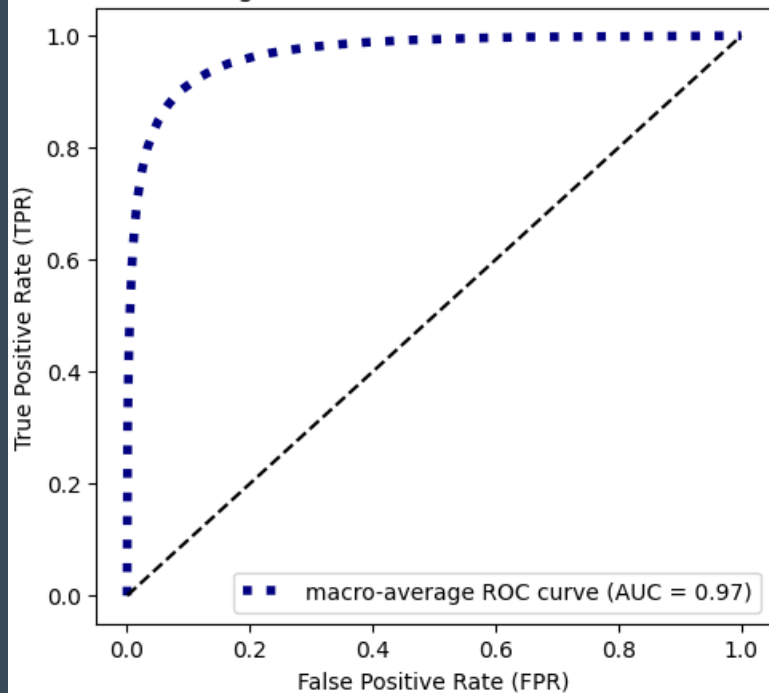
dep_time_cat

- ☐ a) Morning
- ☐ b) Afternoon
- ☒ c) Evening

Weekday



Macro-average ROC Curve - Non-Tuned Random Forest



Macro-average ROC Curve - Tuned Random Forest

