

Anamitra Bhattacharyya
Predict 420-DL, Section 55
Assignment 5 (May 29, 2016)

1) Python Code

```
import pandas as pd #import panda
import numpy as np
#all files are in the current working directory
import cPickle as pickle

#Q1) Import passenger csv file into a pandas DataFrame
passengerDF=pd.read_csv("2014PassengerRawData2-1.csv", sep=',')
len(passengerDF)#25593 records
passengerDF.duplicated().sum()#no entirely duplicated records

#Looked at data for these airports none are a destination airport

sfoDF = passengerDF.loc[passengerDF['OriginApt']=="SFO"]
len(sfoDF)#761 deps from SFO, no instances where SFO is a destination (arrival)!
sfoDF.loc[sfoDF['Total']].sum()#158,535
sfoDF.sort_values(['Total','Carrier'], ascending=False)#AC=largest departure carrier for SFO
sfoCarrier = sfoDF.groupby('Carrier')
sfoResult = sfoCarrier['Total'].aggregate(np.sum)
sfoResult#Best carrier for SFO is UA=3,286,046 total passengers

laxDF = passengerDF.loc[passengerDF['OriginApt']=="LAX"]
len(laxDF)#1456 deps from LAX but none as destination (arrival)!
laxDF.loc[laxDF['Total']].sum()#719,960
laxDF.sort_values(['Total','Carrier'], ascending=False)#AC=largest departure carrier for LAX
laxCarrier = laxDF.groupby('Carrier')
laxResult = laxCarrier['Total'].aggregate(np.sum)
laxResult#Best carrier for LAX is DL total 1,159,667 passengers

atlDF = passengerDF.loc[passengerDF['OriginApt']=="ATL"]
len(atlDF)#1026 deps from ATL, none as destination (arrival)!
atlDF.loc[atlDF['Total']].sum()#501,702
atlDF.sort_values(['Total','Carrier'], ascending=False)#DL=largest departure carrier for ATL
atlCarrier = atlDF.groupby('Carrier')
atlResult = atlCarrier['Total'].aggregate(np.sum)
atlResult#Best carrier for ATL is DL total 8,203,111 passengers

miaDF = passengerDF.loc[passengerDF['OriginApt']=="MIA"]
len(miaDF)#2101 deps from MIA, none as destination (arrival)!
miaDF.loc[miaDF['Total']].sum()#1,039,492
miaDF.sort_values(['Total','Carrier'], ascending=False)#AA=largest departure carrier for MIA
miaCarrier = miaDF.groupby('Carrier')
miaResult = miaCarrier['Total'].aggregate(np.sum)
miaResult#Best carrier for MIA is AA total 11,066,495 passengers

jfkDF = passengerDF.loc[passengerDF['OriginApt']=="JFK"]#2178 deps from JFK, none as destination!
len(jfkDF)#2178 deps from JFK, none as destination (arrival)!
jfkDF.loc[jfkDF['Total']].sum()#1,841,657
jfkDF.sort_values(['Total','Carrier'], ascending=False)#BA=largest departure carrier
jfkCarrier = jfkDF.groupby('Carrier')
jfkResult = jfkCarrier['Total'].aggregate(np.sum)
jfkResult#Best carrier for JFK is B6 total 2,951,853 passengers

#Result: JFK > MIA > LAX > ATL > SFO (passenger departures)
#Result: No arrivals for any of these airports!

#Q2)For each of these airports, determine the airport that the largest number of departures went to.
from pandas import Categorical
sfo_status=Categorical(sfoDF.DestApt)
sfo_status.describe()# ICN and YVR each have most arrivals (50 counts) from SFO

lax_status=Categorical(laxDF.DestApt)
lax_status.describe()# YVR has most arrivals (67 counts) from LAX

atl_status=Categorical(atlDF.DestApt)
```

```

atl_status.describe()# YUL has most arrivals (30 counts) from ATL

mia_status=Categorical(miaDF.DestApt)
mia_status.describe()# SNU has most arrivals (60 counts) from MIA

jfk_status=Categorical(jfkDF.DestApt)
jfk_status.describe()# YYZ has most arrivals (63 counts) from JFK

#Q3)Accidents and deaths
accidentsData=pd.read_table('A2010_14.txt')
len(accidentsData)
sfo1 = accidentsData[accidentsData['c143'].str.strip()=='SFO']
len(sfo1)#4 incidents
# 'c143'=aptCode; 'c6'=Year;'c76'=total fatalities
sfoKilled = sfo1[['c143','c6','c76']]
sfoKilled.columns = ['AptCode','Year','TotFatalities']#3 deaths

lax1 = accidentsData[accidentsData['c143'].str.strip()=='LAX']
len(lax1)#13 incidents
laxKilled = lax1[['c143','c6','c76']]
laxKilled.columns = ['AptCode','Year','TotFatalities']#0 deaths

atl1 = accidentsData[accidentsData['c143'].str.strip()=='ATL']
len(atl1)#24 incidents
atlKilled = atl1[['c143','c6','c76']]
atlKilled.columns = ['AptCode','Year','TotFatalities']#0 deaths

mia1 = accidentsData[accidentsData['c143'].str.strip()=='MIA']
len(mia1)#9 incidents
miaKilled = mia1[['c143','c6','c76']]
miaKilled.columns = ['AptCode','Year','TotFatalities']#0 deaths

jfk1 = accidentsData[accidentsData['c143'].str.strip()=='JFK']
len(jfk1)#10 incidents
jfkKilled = jfk1[['c143','c6','c76']]
jfkKilled.columns = ['AptCode','Year','TotFatalities']#0 deaths

#Q4)More on fatalities...
allKilled = accidentsData[['c143','c6','c76','c77']]#extract fatality data to a new df
allKilled.columns = ['AptCode','Year','TotFatalities','PrimaryCause']
len(allKilled)#11501 entries

fatalitiesDF = allKilled.loc[allKilled['TotFatalities'] > 0.0]#extract incidents with >= 1 fatality
len(fatalitiesDF)#1096 instances of > 0 deaths
fatalitiesDF.dropna(subset=['PrimaryCause'], inplace=True)#remove NaN instances (15)
len(fatalitiesDF)#1081
sortFatalities = fatalitiesDF.sort_values(['TotFatalities','PrimaryCause'], ascending=False)
#Couldn't find way to remove empty cells so exported to Excel to do it:
sortFatalities.to_csv("Fatalities.csv")
sortFilterFatalities=pd.read_csv("Fatalities-filtered.csv", sep=',')
len(sortFilterFatalities)#41 instances sorted by freq of occurrence of fatalities
sortFilterFatalities[0:13]#to display top 10 different ATA codes sorted by no. of fatalities

```

Example output:

Each airlines departures/arrivals carrier

In [1037]: sfoDF.sort_values(['Total','Carrier'], ascending=False)

Out[1037]:

	Month	OriginApt	DestApt	OriginWAC	DestWAC	Carrier	Group	Type	Total	Scheduled	Charter
15100	201407	SFO	YYZ	91	936	AC	0	Passengers	48917	48,917	0
17322	201408	SFO	YYZ	91	936	AC	0	Passengers	48137	48,137	0
12801	201406	SFO	YYZ	91	936	AC	0	Passengers	44035	44,035	0

In [1038]: laxDF.sort_values(['Total','Carrier'], ascending=False)

Out[1038]:

	Month	OriginApt	DestApt	OriginWAC	DestWAC	Carrier	Group	Type	Total	Scheduled	Charter
14294	201407	LAX	YYZ	91	936	AC	0	Passengers	62504	62,504	0
16495	201408	LAX	HKG	91	729	CX	0	Passengers	61962	61,962	0
16579	201408	LAX	YYZ	91	936	AC	0	Passengers	61410	61,410	0

In [1039]: atlDF.sort_values(['Total','Carrier'], ascending=False)

Out[1039]:

	Month	OriginApt	DestApt	OriginWAC	DestWAC	Carrier	Group	Type	Total	Scheduled	Charter
4291	201403	ATL	CUN	34	148	DL	3	Passengers	55480	55,480	0
12999	201407	ATL	CUN	34	148	DL	3	Passengers	55088	55,088	0
10765	201406	ATL	CUN	34	148	DL	3	Passengers	51757	51,757	0

In [1041]: miaDF.sort_values(['Total','Carrier'], ascending=False)

Out[1041]:

	Month	OriginApt	DestApt	OriginWAC	DestWAC	Carrier	Group	Type	Total	Scheduled	Charter
1433	201401	MIA	GRU	33	316	AA	3	Passengers	51349	51,349	0
24877	201412	MIA	GRU	33	316	JJ	0	Passengers	51198	51,198	0
24941	201412	MIA	PTY	33	162	CM	0	Passengers	48637	48,637	0

In [1042]: jfkDF.sort_values(['Total','Carrier'], ascending=False)

Out[1042]:

	Month	OriginApt	DestApt	OriginWAC	DestWAC	Carrier	Group	Type	Total	Scheduled	Charter
16313	201408	JFK	LHR	22	493	BA	0	Passengers	134263	134,263	0
20417	201410	JFK	LHR	22	493	BA	0	Passengers	132822	132,822	0
18438	201409	JFK	LHR	22	493	BA	0	Passengers	132052	132,052	0

Accidents/deaths (output)

SFO: 3 deaths from 4 accidents/incidents

In [1203]: sfoKilled

Out[1203]:

	AptCode	Year	TotFatalities
1721	SFO	2010	0.0
4208	SFO	2011	0.0
5324	SFO	2011	0.0
10842	SFO	2013	3.0

LAX: 0 deaths from 13 accidents/incidents

In [1193]: laxKilled

Out[1193]:

	AptCode	Year	TotFatalities
312	LAX	2010	0.0
1475	LAX	2010	0.0
2151	LAX	2010	0.0
2498	LAX	2010	0.0
2857	LAX	2011	0.0
3847	LAX	2011	0.0
5264	LAX	2011	0.0
5615	LAX	2011	0.0
5968	LAX	2012	0.0
6450	LAX	2012	0.0
7764	LAX	2012	0.0
9358	LAX	2013	0.0
10495	LAX	2013	0.0

ATL: 0 deaths from 24 incidents (but some data missing)

```
In [1196]: atkKilled
```

```
Out[1196]:
```

	AptCode	Year	TotFatalities
7	ATL	2010	0.0
25	ATL	2010	0.0
42	ATL	2010	0.0
95	ATL	2010	0.0
152	ATL	2010	0.0
159	ATL	2010	0.0
173	ATL	2010	0.0
210	ATL	2010	0.0
220	ATL	2010	0.0
223	ATL	2010	0.0
242	ATL	2010	0.0
245	ATL	2010	0.0
283	ATL	2010	0.0
458	ATL	2010	0.0
501	ATL	2010	0.0
847	ATL	2010	0.0
848	ATL	2010	0.0
2631	ATL	2010	0.0
3948	ATL	2011	0.0
4073	ATL	2011	0.0
4257	ATL	2011	0.0
6177	ATL	2012	0.0
9636	ATL	2013	0.0
10459	ATL	2013	0.0

MIA: 0 deaths from 9 incidents (some data missing)

In [1198]: miaKilled

Out[1198]:

	AptCode	Year	TotFatalities
43	MIA	2010	0.0
875	MIA	2010	0.0
2367	MIA	2010	0.0
2466	MIA	2010	0.0
2597	MIA	2010	0.0
3433	MIA	2011	0.0
3517	MIA	2011	0.0
4158	MIA	2011	0.0
8552	MIA	2012	0.0

JFK: 0 deaths from 10 incidents (some data missing)

In [1200]: jfkKilled

Out[1200]:

	AptCode	Year	TotFatalities
1557	JFK	2010	0.0
1848	JFK	2010	0.0
2081	JFK	2010	0.0
6538	JFK	2012	0.0
7777	JFK	2012	0.0
8819	JFK	2013	0.0
8927	JFK	2012	0.0
9037	JFK	2012	0.0
9649	JFK	2013	0.0
10646	JFK	2013	0.0

Incident codes sorted and filtered:

In [1282]: sortFilterFatalities[0:13]

Out[1282]:

	AptCode	Year	TotFatalities	PrimaryCause
0	K4SD	2011	12	5552
1	BLM	2010	5	5700
2	LCH	2013	3	6410
3	KBDU	2010	3	6114
4	X59	2011	2	8530
5	NaN	2010	2	8500
6	NaN	2010	2	8500
7	NaN	2010	2	8500
8	NaN	2012	2	8500
9	PWM	2010	2	7314
10	U42	2011	2	7300
11	NaN	2013	2	7261
12	SDP	2010	2	7200

These are the descriptors for FAA incident codes for the top 10 incident codes for this data sorted by no. of fatalities, that I could find:

Total Fatalities	Incident code	Primary Cause (ATA Code)
12	5552	
5	5700	
3	6410	
3	6114	
2	8530	Reciprocating Engine Cylinder section
2	8500	Engine (reciprocating)
2	8500	Engine (reciprocating)
2	8500	Engine (reciprocating)
2	8500	Engine (reciprocating)
2	7314	
2	7300	
2	7261	
2	7200	