```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('flights.csv')
df = df.dropna()

num = df._get_numeric_data()

df.loc[df['arr_delay']<0,'arr_delay']=0
df.loc[df['dep_delay']<0,'dep_delay']=0

#df['dep_delay'].clip(lower=0)
num.head((6))
```

> /usr/local/lib/python3.7/dist-packages/pandas/core/indexing.py:1817: SettingWith
> A value is trying to be set on a copy of a slice from a DataFrame.
> Try using .loc[row_indexer,col_indexer] = value instead
>
> See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stab
>   self._setitem_single_column(loc, value, pi)

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	flight	air_time
0	2013	1	1	517.0	2.0	830.0	11.0	1545	227.0
1	2013	1	1	533.0	4.0	850.0	20.0	1714	227.0
2	2013	1	1	542.0	2.0	923.0	33.0	1141	160.0
3	2013	1	1	554.0	0.0	812.0	0.0	461	116.0
4	2013	1	1	554.0	0.0	740.0	12.0	1696	150.0
5	2013	1	1	558.0	0.0	753.0	8.0	301	138.0

```
len(df)
```

> 157927

Question 1: 1) (25 pts) What is the average departure delay time per month and which month did we have the most departure delays overall?

```
#Avg departure delay time per month

df2 = df.groupby('month')['dep_delay'].mean()
df2.head(12)

#July has the most delays by amount of time (19.57)
```

```
month
1      8.776559
2      9.075380
3     12.063348
4     13.322943
5     12.333778
6     19.570434
7     19.222198
8     12.102919
9      8.920526
10     7.309489
11     7.067419
12    14.752218
Name: dep_delay, dtype: float64
```

July has the most delays, 19.57 avg.

2) (25 pts) What is the average arrival and departure delay per air carrier?


```
# Arrival delays:
arrDelay = df.groupby('carrier')['arr_delay'].mean()
arrDelay.head()
```

```
carrier
AA    12.823458
AS     9.160790
DL    12.998552
UA    14.095358
US    10.751147
Name: arr_delay, dtype: float64
```

```
# Departure Delays:
departDelay = df.groupby('carrier')['dep_delay'].mean()
departDelay.head()
```

```
carrier
AA    11.754719
AS     9.988717
DL    11.896429
UA    13.984615
```

```
  US      7.904695
Name: dep_delay, dtype: float64
```

3) (25 pts) What were the top 5 flights with the worst departure delays?

```
worstDelays=df.sort_values(by='dep_delay', ascending=False)
worstDelays.head(5)
```

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailn
156166	2013	9	20	1139.0	1014.0	1457.0	1007.0	AA	N338
82913	2013	4	10	1100.0	960.0	1342.0	931.0	DL	N959
72379	2013	3	17	2321.0	911.0	135.0	915.0	DL	N927
117991	2013	6	27	959.0	899.0	1236.0	850.0	DL	N376
129429	2013	7	22	2257.0	898.0	121.0	895.0	DL	N671

(25 pts) How much delay do I need to consider when I use AA in July? (based on average)

```
aaJuly=df[(df['month']==7) & (df['carrier']== 'AA')]
#aaJuly.head()

# departure delay
dep_del=aaJuly['dep_delay'].mean()

# answ: 13.934155

# arribal delay
arr_del=aaJuly['arr_delay'].mean()

# answ: 16.693029

dep_del+arr_del
```

```
   31.567713976164683
```

Average consider about 13.934 mins departure delay, and 16.693 mins arrival delay, so 31.57 minutes total.

Colab paid products - Cancel contracts here

✓ 0s   completed at 7:09 PM