Code git conda_amazonei_mxnet_p36

[ ]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[ ]:
```
df = pd.read_csv("train.csv")
```

[3]:
```
df.shape
```

[3]:
```
(71549, 25)
```

[12]:
```
df.shape
```

[12]:
```
(71549, 25)
```

[10]:
```
df.dropna()
```

[10]:

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	...	Inflight entertainment	On-board service	Leg room service	Bag han
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	...	5	4	3	
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	...	1	1	5	
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	...	5	4	3	
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	...	2	2	5	
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	...	3	3	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
71543	71543	67255	Female	Loyal Customer	34	Personal Travel	Eco	1119	2	5	...	5	3	5	
71544	71544	70989	Male	Loyal Customer	45	Business travel	Business	3824	5	5	...	4	4	4	
71545	71545	122454	Male	Loyal Customer	45	Business travel	Business	2162	2	2	...	4	4	4	

[14]:
```
df.duplicated()
```

[14]:
```
0        False
1        False
2        False
3        False
4        False
         ...
71544    False
71545    False
71546    False
71547    False
71548    False
Length: 71549, dtype: bool
```

[15]:
```
df.duplicated().sum()
```

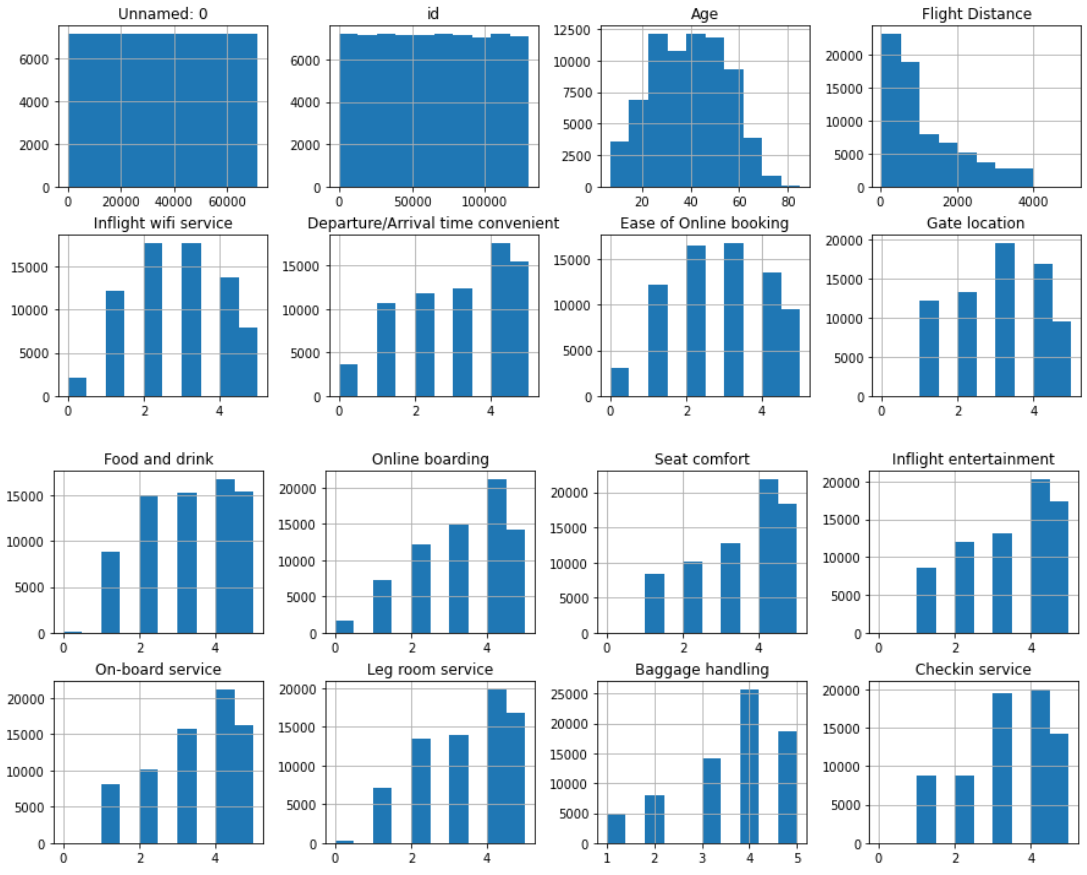
[15]:
```
0
```

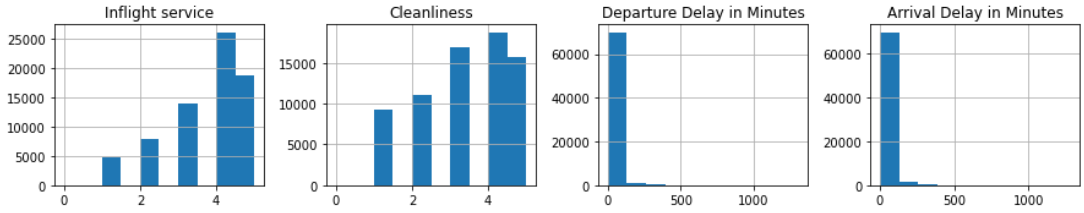
[16]:
```
df.describe().T
```

[16]:

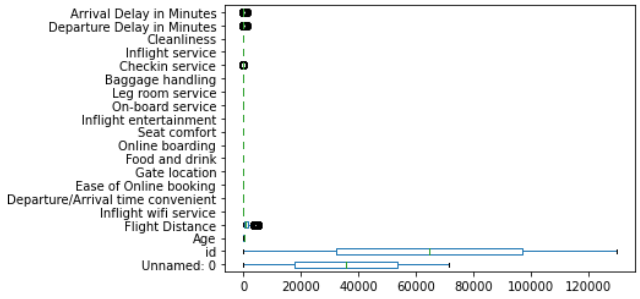
	count	mean	std	min	25%	50%	75%	max
Unnamed: 0	71549.0	35774.000000	20654.561542	0.0	17887.0	35774.0	53661.0	71548.0
id	71549.0	64815.304840	37475.852677	2.0	32406.0	64714.0	97240.0	129880.0
Age	71549.0	39.382857	15.099017	7.0	27.0	40.0	51.0	85.0
Flight Distance	71549.0	1187.926722	996.140346	31.0	413.0	840.0	1739.0	4983.0
Inflight wifi service	71549.0	2.733204	1.330131	0.0	2.0	3.0	4.0	5.0
Departure/Arrival time convenient	71549.0	3.059735	1.527065	0.0	2.0	3.0	4.0	5.0
Ease of Online booking	71549.0	2.756391	1.399490	0.0	2.0	3.0	4.0	5.0
Gate location	71549.0	2.973934	1.278876	0.0	2.0	3.0	4.0	5.0
Food and drink	71549.0	3.205565	1.331446	0.0	2.0	3.0	4.0	5.0
Online boarding	71549.0	3.252359	1.349354	0.0	2.0	3.0	4.0	5.0
Seat comfort	71549.0	3.442941	1.320747	0.0	2.0	4.0	5.0	5.0
Inflight entertainment	71549.0	3.359027	1.333851	0.0	2.0	4.0	4.0	5.0
On-board service	71549.0	3.381669	1.288191	0.0	2.0	4.0	4.0	5.0
Leg room service	71549.0	3.348279	1.312775	0.0	2.0	4.0	4.0	5.0

[19]:
```
hist = df.hist(bins=10,figsize =(15,15))
```





[23]:
```
bxplt = df.boxplot(grid=False, vert=False,fontsize=10)
```



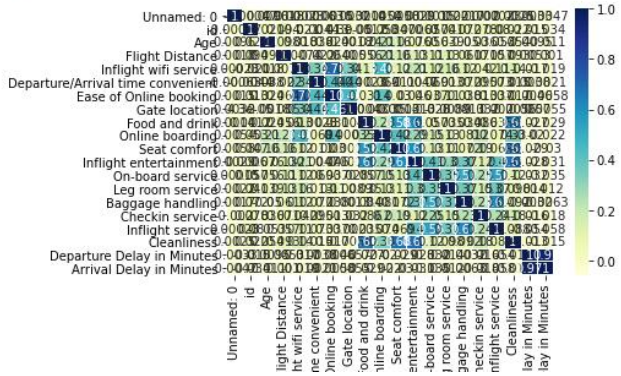
[24]:
```
plt.figure(figsize=(14,12))
```

[24]:
```
<Figure size 1008x864 with 0 Axes>
```

```
<Figure size 1008x864 with 0 Axes>
```


[27]:
```
sns.heatmap(df.corr(),linewidths=.1,cmap="YlGnBu", annot=True)
```

[27]:
```
<AxesSubplot:>
```



[29]:
```
g = sns.pairplot(df)
```