

```
In [29]: import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
df=pd.read_csv('Dataset (6).csv',sep=',')
df.head()
```

	Uniq Id	Package Name	Package Type	Destination	Itinerary	Places Covered	Travel Date	Hotel Details	Start City	Airline	Flight Stops	Meals	Sightseeing Places Covered	Cancellation
0	e78ba76d9b8d1e6ed2139645ca5d1	Best of Shimla and Manali Holiday from Delhi	Standard	New Delhi(Shimla)Manali(Chandigarh)	1N New Delhi, 2N Shimla, 2N Manali, 1N Chandi...	New Delhi(Shimla)Manali(Chandigarh)	30-07-2021	Not Available	Mumbai	Not Available	2	3	Not Available	Not Avail...
1	1788982630c3c335a5a41d5b39370	Kashmir Valley vacation	Premium	Srinagar(Pahalgam)Srinagar	1N Srinagar, 2N Pahalgam, 1N Srinagar	Srinagar(Pahalgam)Srinagar	09-12-2021	The Orchard Retreat & Spa,4 @Welcomhotel Pinc...	New Delhi	IndiGo IndiGo	0	5	Dal Lake Awaritara Purni Mughar Gardens ...	Cancellation after in the 1st
2	k0602954d40502c2fda935b7b7df	Might of Mewar- Udaipur and Chitorgarh	Luxury	Udaipur(Chitorgarh)	2N Udaipur, 1N Chitorgarh	Udaipur(Chitorgarh)	26-04-2021	The Anamas 4 @Ji5Tha Lake Nahargarh Palace-4	New Delhi	IndiGo	0	4	Lake Pichola Jag Mandir Palace Sahelyan ...	Cancellation after in the 1st
3	32a19a6c171e674482346da46d193c	Colorful Kerala (Romantic Getaway)	Premium	Munnar(Kumarakom)Alleppey(Kovalam and Poovar	2N Munnar, 1N Kumarakom, 1N Alleppey, 2N Koval...	Munnar(Kumarakom)Alleppey(Kovalam and Poovar	27-08-2021	Elkor Hills Suites Resort & Spa-MMT Holidays ...	New Delhi	IndiGo	0	5	Mattupetty Dam Echo Point Tea Museum...	Cancellation after in the 1st
4	1076068aada0a3bc25d24869594d109f	A Week in Bangkok & Pattaya	Premium	Pattaya(Bangkok	4N Pattaya, 2N Bangkok	Pattaya(Bangkok	12-12-2021	Dusit Thani Pattaya - MMT Special 4 @Manat Villa...	New Delhi	SpiceJet Go Air	0	5	Coral Island Tour with Indian Lunch, 3N Spa...	Cancellation after in the 1st

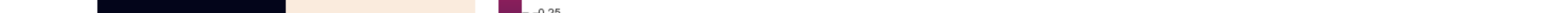
```
In [30]: codf=df.select_dtypes(include='object')
codf
```

	Uniq Id	Package Name	Package Type	Destination	Itinerary	Places Covered	Travel Date	Hotel Details	Start City	Airline	Flight Stops	Meals	Sightseeing Places Covered	Cancellation
0	e78ba76d9b8d1e6ed2139645ca5d1	Best of Shimla and Manali Holiday from Delhi	Standard	New Delhi(Shimla)Manali(Chandigarh)	1N New Delhi, 2N Shimla, 2N Manali, 1N Chandi...	New Delhi(Shimla)Manali(Chandigarh)	30-07-2021	Not Available	Mumbai	Not Available	2	3	Not Available	Not Avail...
1	1788982630c3c335a5a41d5b39370	Kashmir Valley vacation	Premium	Srinagar(Pahalgam)Srinagar	1N Srinagar, 2N Pahalgam, 1N Srinagar	Srinagar(Pahalgam)Srinagar	08-12-2021	The Orchard Retreat & Spa,4 @Welcomhotel Pinc...	New Delhi	IndiGo IndiGo	0	5	Dal Lake Awaritara Purni Mughar Gardens ...	Cancellation after in the 1st
2	k0602954d40502c2fda935b7b7df	Might of Mewar- Udaipur and Chitorgarh	Luxury	Udaipur(Chitorgarh)	2N Udaipur, 1N Chitorgarh	Udaipur(Chitorgarh)	26-04-2021	The Anamas 4 @Ji5Tha Lake Nahargarh Palace-4	New Delhi	IndiGo	0	4	Lake Pichola Jag Mandir Palace Sahelyan ...	Cancellation after in the 1st
3	32a19a6c171e674482346da46d193c	Colorful Kerala (Romantic Getaway)	Premium	Munnar(Kumarakom)Alleppey(Kovalam and Poovar	2N Munnar, 1N Kumarakom, 1N Alleppey, 2N Koval...	Munnar(Kumarakom)Alleppey(Kovalam and Poovar	27-08-2021	Elkor Hills Suites Resort & Spa-MMT Holidays ...	New Delhi	IndiGo	0	5	Mattupetty Dam Echo Point Tea Museum...	Cancellation after in the 1st
4	1076068aada0a3bc25d24869594d109f	A Week in Bangkok & Pattaya	Premium	Pattaya(Bangkok	4N Pattaya, 2N Bangkok	Pattaya(Bangkok	12-12-2021	Dusit Thani Pattaya - MMT Special 4 @Manat Villa...	New Delhi	SpiceJet Go Air	0	5	Coral Island Tour with Indian Lunch, 3N Spa...	Cancellation after in the 1st
...
20995	ae3ae6ca299180470145d6154d2f2b	Shimla & Manali Retreat from Delhi	Deluxe	Shimla(Manali)New Delhi	2N Shimla, 3N Manali, 1N New Delhi	Shimla(Manali)New Delhi	15-04-2021	Summit Thistle Villa Resort & Spa,Four @Snow Va...	Mumbai	Not Available			Mall road Indira Banglow Himalayan Zoo ...	Cancellation after in the 1st
20996	6dc553b63597600909092bea2be5e5	Bedazzling Kashmir Holiday	Luxury	Srinagar(Gumarg)Pahalgam(Srinagar	1N Srinagar, 2N Gumarg, 1N Pahalgam, 3N Srin...	Srinagar(Gumarg)Pahalgam(Srinagar	30-08-2021	The Orchard Retreat & Spa,4 @The Rhyber Hotel...	New Delhi	Vistara Vistara			Dal Lake Awaritara Purni Mughar Gardens ...	Cancellation after in the 1st
20997	d3ba12f6949c5502320242c45b676007	A relaxing week in (Berata and Colib)	Deluxe	Colombo	5N Colombo	Colombo	27-02-2021	Mount Lavinia Hotel HOLIDAYS SPECIAL 4.0	New Delhi	IndiGo IndiGo IndiGo IndiGo			Mall road Indira Banglow Himalayan Zoo ...	Cancellation after in the 1st
20998	51d843b6953ec7ab6c7299431111b05	Kerala North East with Pelling and Cochin	Deluxe	Darjeeling(Pelling)Gangtok	2N Darjeeling, 2N Pelling, 2N Gangtok	Darjeeling(Pelling)Gangtok	13-01-2022	Ramada by Wyndham Darjeeling Gandhi Road,4 @T...	New Delhi	IndiGo			Tiger Hill Himalayan Rues Mughar Gardens ...	Cancellation after in the 1st
20999	5ed5968f7aba743f92a269384c04d54	Munnar, Kovalam, Alleppey Holiday	Standard	Munnar(Kovalam)Alleppey	2N Munnar, 2N Kovalam, 1N Alleppey	Munnar(Kovalam)Alleppey	20-05-2021	Isoberg Hill Hotel- MMT Holidays Special 3.8 @Sa...	New Delhi	IndiGo			Tea Tea Museum Eravikulam national park ...	Cancellation after in the 1st

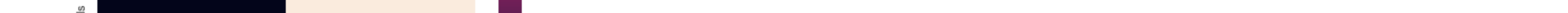
21000 rows x 12 columns

We can observe that premium and Luxury type packages have high prices per person

```
In [31]: sns.set(rc={'figure.figsize':(8,8)})
sns.barplot(data=df,x='Package Type',y='Per Person Price')
```



```
In [32]: sns.barplot(data=df,x='Start City',y='Per Person Price')
```



correlation of independent variables in a Correlation Matrix

```
In [33]: codf=df.drop(['Per Person Price'],axis=1)
ax = sns.heatmap(codf.corr(), annot=True)
```



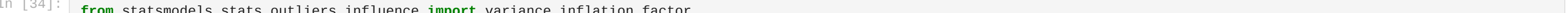
VIF values

```
In [34]: from statsmodels.stats.outliers_influence import variance_inflation_factor
def calc_vif(X):
    # Calculating VIF
    vif = pd.DataFrame()
    vif['variables'] = X.columns
    vif['vif'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
    return vif
```

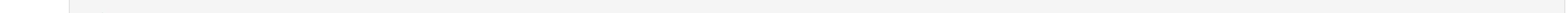
```
In [35]: X=df.iloc[:,10:12]
calc_vif(X)
```

variables	VIF
0 Flight Stops	1.466572
1 Meals	1.466572

```
In [36]: sns.histplot(data=df,x='Flight Stops',y='Per Person Price',hue='Start City')
```



```
In [37]: sns.scatterplot(data=df,x='Meals',y='Per Person Price',hue='Package Type')
```



```
In [38]: ls=[]
ls0=[]
for i in df.columns:
    if df['{}'].nunique()<11:
        ls.append(i)
        ls0.append(i)
ls,ls0
```

Out[38]: ([5, 2, 3, 4, 38], ['Package Type', 'Start City', 'Flight Stops', 'Meals', 'Cancellation Rules'])

```
In [39]: ls='rm'
print('{}'.format(i))
```

rm

```
In [40]: Cancellation_Rules = df['Cancellation Rules'].unique()
for i, j in enumerate(Cancellation_Rules):
    df.replace(j,i,inplace=True)
pd.crosstab(df['Cancellation Rules'],df['Package Type'])
df['Cancellation Rules'].value_counts()
```

1	17574
2	2616
3	795
4	7
5	2
6	7
7	2
8	1
9	1
10	1
11	1

Name: Cancellation Rules, dtype: int64

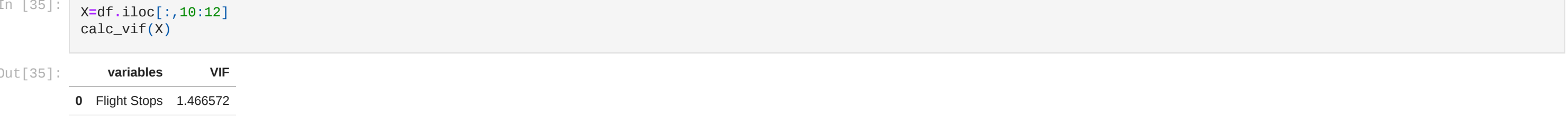
```
In [41]: df['Package Type'].value_counts()
```

Deluxe	6086
Standard	5669
Premium	5581
Luxury	3278
Budget	466

Name: Package Type, dtype: int64

Pie chart of Package type displaying the value's frequency distribution

```
In [79]: plt.figure(figsize=(10,10))
plt.pie(df['Package Type'].value_counts(),labels=df['Package Type'].unique(),autopct='%1.1f%%')
plt.legend(loc='upper left')
plt.show()
```



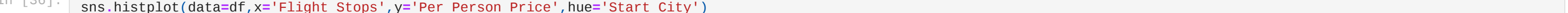
```
In [198]: df['year'] = pd.DatetimeIndex(df['Travel Date']).year
df['month'] = pd.DatetimeIndex(df['Travel Date']).month
df['places'].value_counts()
```

5	2689
7	1051
4	1941
6	1940
3	1934
9	1845
2	1816
10	1880
8	1673
11	1621
12	1391
1	1169

Name: month, dtype: int64

Year Wise-Package type value's frequency distribution

```
In [200]: sns.countplot(x='year', hue = 'Package Type', data = df,
dodge = True)
plt.xlabel(xlabel='No of Places Visited')
plt.xticks(rotation = 45, ha = 'right')
plt.legend(title = 'Package Type', bbox_to_anchor = (1, 1))
```



```
In [169]: df['Package Type'].value_counts()
```

Deluxe	6086
Standard	5669
Premium	5581
Luxury	3278
Budget	466

Name: Package Type, dtype: int64

```
In [132]: x=[]
for i in df['destination']:
    bi.split(':')[1]
    a.append(len(b))
df['places']=a
df['places'].value_counts()
```

3	5385
1	5383
2	5332
4	3363
5	1849
6	177
7	182
8	31
9	12
10	6
11	6

No of places visited-Package type value's frequency distribution

```
In [209]: sns.countplot(x='places', hue = 'Package Type', data = df,
dodge = True)
plt.xlabel(xlabel='No of Places Visited')
plt.xticks(rotation = 45, ha = 'right')
plt.legend(title = 'Package Type', bbox_to_anchor = (1, 1))
```



Start City-Package type value's frequency distribution

```
In [214]: sns.histplot(data=df,x='Start City', hue='Package Type', multiple='dodge', shrink=1)
```



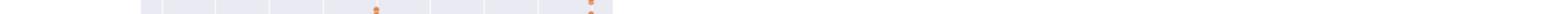
```
In [215]: df['year'] = pd.DatetimeIndex(df['Travel Date']).year
df['month'] = pd.DatetimeIndex(df['Travel Date']).month
df['places'].value_counts()
```

5	2689
7	1051
4	1941
6	1940
3	1934
9	1845
2	1816
10	1880
8	1673
11	1621
12	1391
1	1169

Name: month, dtype: int64

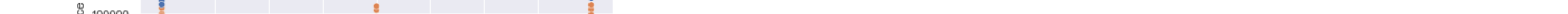
Year Wise-Package type value's frequency distribution

```
In [220]: sns.countplot(x='places', hue = 'Package Type', data = df,
dodge = True)
plt.xlabel(xlabel='No of Places Visited')
plt.xticks(rotation = 45, ha = 'right')
plt.legend(title = 'Package Type', bbox_to_anchor = (1, 1))
```



Bivariate Plot of No of Places visited- Package Type values with respect to the Per Person Price

```
In [225]: sns.barplot(data=df,x='places',y='Per Person Price',hue='Package Type')
```



Correlation matrix of all independent variables

```
In [226]: codf=df.drop(['Per Person Price'],axis=1)
ax = sns.heatmap(codf.corr(), annot=True)
```



VIF for independent variables

```
In [227]: X=df.iloc[:,(10,11,15,16,17)]
calc_vif(X)
```

variables	VIF
0 Flight Stops	6.922562
1 Meals	59.155843
2 year	90.579966
3 month	5.239632
4 places	4.899320