```
# Import python libraries
# Pandas and numpy
import pandas as pd
import numpy as np
# Matplotlib and seaborn for plots and % visualizing data
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns

# Reading CSV file
df = pd.read_csv("C:/Users/admin/Desktop/Online%20Retail/Online%20Retail.csv", encoding = 'unicode_es')

# Versions used for of solver and read csv
print(pd.__version__)
1.0.5

# Checking top 5 rows of dataset
df.head()
```

	User ID	Year	Month	Day	Hour	Minute	Second	Age	Gender	Marital Status	State	Zone	Occupation	Product Category	Orders	Amount	Status	unnamed0
0	1000001	Female	P00125942	7	17	15.0	36	26-35	F	Married	Maharashtra	Western	Healthcare	Auto	1	15035.0	Valid	Valid
1	1000002	Male	P00110742	3	12	19.0	35	51-55	M	Unmarried	Karnataka	Southern	Construction	Auto	2	23233.0	Valid	Valid
2	1001008	Male	P00117942	10	12	07.0	39	36-45	M	Married	Madhya Pradesh	Southern	Auto	Auto	3	20651.0	Valid	Valid
3	1001009	Male	P00010742	11	12	19.0	41	46-50	M	Married	Karnataka	Southern	Construction	Auto	4	20372.0	Valid	Valid
4	1003004	Male	P00026742	7	12	19.0	36	26-35	M	Married	Gujarat	Western	Food Processing	Auto	5	19214.0	Valid	Valid

```
df.info()
```

	User ID	Year	Month	Day	Hour	Minute	Second	Age	Gender	Marital Status	State	Zone	Occupation	Product Category	Orders	Amount	Status	unnamed0
0	1000001	Female	P00125942	7	17	15.0	36	26-35	F	Married	Maharashtra	Western	Healthcare	Auto	1	15035.0	Valid	Valid
1	1000002	Male	P00110742	3	12	19.0	35	51-55	M	Unmarried	Karnataka	Southern	Construction	Auto	2	23233.0	Valid	Valid
2	1001008	Male	P00117942	10	12	07.0	39	36-45	M	Married	Madhya Pradesh	Southern	Auto	Auto	3	20651.0	Valid	Valid
3	1001009	Male	P00010742	11	12	19.0	41	46-50	M	Married	Karnataka	Southern	Construction	Auto	4	20372.0	Valid	Valid
4	1003004	Male	P00026742	7	12	19.0	36	26-35	M	Married	Gujarat	Western	Food Processing	Auto	5	19214.0	Valid	Valid

```
# Checking shape of dataset
df.shape
(11251, 18)

# Checking top 5 rows for dataset
df.head()
```

	User ID	Year	Month	Day	Hour	Minute	Second	Age	Gender	Marital Status	State	Zone	Occupation	Product Category	Orders	Amount	Status	unnamed0
0	1000001	Female	P00125942	7	17	15.0	36	26-35	F	Married	Maharashtra	Western	Healthcare	Auto	1	15035.0	Valid	Valid
1	1000002	Male	P00110742	3	12	19.0	35	51-55	M	Unmarried	Karnataka	Southern	Construction	Auto	2	23233.0	Valid	Valid
2	1001008	Male	P00117942	10	12	07.0	39	36-45	M	Married	Madhya Pradesh	Southern	Auto	Auto	3	20651.0	Valid	Valid
3	1001009	Male	P00010742	11	12	19.0	41	46-50	M	Married	Karnataka	Southern	Construction	Auto	4	20372.0	Valid	Valid
4	1003004	Male	P00026742	7	12	19.0	36	26-35	M	Married	Gujarat	Western	Food Processing	Auto	5	19214.0	Valid	Valid

```
# Checking bottom 5 rows for dataset
df.tail()
```

	User ID	Year	Month	Day	Hour	Minute	Second	Age	Gender	Marital Status	State	Zone	Occupation	Product Category	Orders	Amount	Status	unnamed0
11246	1000995	Female	P00265242	2	17	01.0	35	51-55	F	Married	Karnataka	Southern	Chemical	Beauty	3	370.0	Valid	Valid
11247	1001265	Male	P00184942	7	18	35.0	35	Unknown	M	Married	Madhya Pradesh	Southern	Food	Auto	2	20604	Valid	Valid
11248	1001363	Male	P00220442	6	18	35.0	55	26-35	M	Married	Uttar Pradesh	Central	Healthcare	Auto	1	7197.0	Valid	Valid
11249	1001375	Male	P00058442	8	18	35.0	26	Unknown	M	Married	Karnataka	Southern	Construction	Auto	2	19842	Valid	Valid
11250	1001918	Female	P00039342	6	19	41.0	35	Unknown	F	Married	Maharashtra	Western	Agriculture	Office	2	368	Valid	Valid
11251	1002035	Female	P00059442	10	20	43.0	36	Unknown	F	Married	Maharashtra	Western	Healthcare	Office	3	188	Valid	Valid

```
11251rows x 18 columns

# Checking total number of null values
df.isnull().sum()
User ID             0
Year                0
Month               0
Day                 0
Hour                0
Minute              0
Second              0
Age                 0
Gender              0
Marital Status      0
State               0
Zone                0
Occupation          0
Product Category    0
Orders              0
Amount              0
Status              0
unnamed0            0
dtype: int64

# Checking data types
df.dtypes
User ID             int64
Year                int64
Month               int64
Day                 int64
Hour                int64
Minute              int64
Second              int64
Age                 int64
Gender              int64
Marital Status      int64
State               int64
Zone                int64
Occupation          int64
Product Category    int64
Orders              int64
Amount              int64
Status              int64
unnamed0            int64
dtype: int64

# Checking unique values
df.nunique()
User ID             1000
Year                   2
Month                  3
Day                   31
Hour                  24
Minute                60
Second                60
Age                    7
Gender                 2
Marital Status         2
State                 14
Zone                   5
Occupation            15
Product Category      20
Orders                21
Amount             9475
Status                 1
unnamed0               1
dtype: int64

# Checking duplicate values
df.duplicated().sum()
0

# Checking for null values
df.isnull().sum()
User ID             0
Year                0
Month               0
Day                 0
Hour                0
Minute              0
Second              0
Age                 0
Gender              0
Marital Status      0
State               0
Zone                0
Occupation          0
Product Category    0
Orders              0
Amount              0
Status              0
unnamed0            0
dtype: int64

# Checking data types
df.dtypes
User ID             int64
Year                int64
Month               int64
Day                 int64
Hour                int64
Minute              int64
Second              int64
Age                 int64
Gender              int64
Marital Status      int64
State               int64
Zone                int64
Occupation          int64
Product Category    int64
Orders              int64
Amount              int64
Status              int64
unnamed0            int64
dtype: int64

# Checking shape
df.shape
(11251, 18)

# Checking data types
df.dtypes
User ID             int64
Year                int64
Month               int64
Day                 int64
Hour                int64
Minute              int64
Second              int64
Age                 int64
Gender              int64
Marital Status      int64
State               int64
Zone                int64
Occupation          int64
Product Category    int64
Orders              int64
Amount              int64
Status              int64
unnamed0            int64
dtype: int64

# Checking unique values
df.nunique()
User ID             1000
Year                   2
Month                  3
Day                   31
Hour                  24
Minute                60
Second                60
Age                    7
Gender                 2
Marital Status         2
State                 14
Zone                   5
Occupation            15
Product Category      20
Orders                21
Amount             9475
Status                 1
unnamed0               1
dtype: int64

# Checking duplicate values
df.duplicated().sum()
0

# Checking for null values
df.isnull().sum()
User ID             0
Year                0
Month               0
Day                 0
Hour                0
Minute              0
Second              0
Age                 0
Gender              0
Marital Status      0
State               0
Zone                0
Occupation          0
Product Category    0
Orders              0
Amount              0
Status              0
unnamed0            0
dtype: int64
```