ANALYZING AND VISUALIZING THE SALES DATA OF A COMPANY USING PYTHON

Mohammed Adil Adnan

Abstract

Now a day's analyzing the data and visualizing is most important criteria to predict the sale of a product in a company. In order to analysis the data we use python in this project. In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Pycharm, or GTK+. In our project we are using Pycharm to implement the both analysis and visualization of the sales data. We are performing some serious statistical function to develop the data Mean, Summation, Maximum, Minimum, Count, Median, Standard deviation, Variance.

INTRODUCTION

Python is a general purpose programming language. Hence, we can use the programming language for developing both desktop and web applications. Also, we can use Python for developing complex scientific and numeric applications. Python is designed with features to facilitate data analysis and visualization.

A. Pandas

In order to work with the data in Python, we'll need to read the .csv file into Pandas DataFrame. A DataFrame is a way to represent and work with tabular data. Tabular data has rows and columns, just like csv file. In order to read in the data, we'll need to use the pandas.read_csv function.

B. Matplotlib

In order to be able to plot the graph or any visualization with the data in Python, we'll need to read the csv file into Matplotlib pyplot. A pyplot is a way to represent and work with tabular data.

EXPERIMENT DESCRIPTION AND METHODS

Table 1:

Year	Customers	Year	Customers
2000	151315	2010	207499
2001	202307	2011	112238
2002	37392	2012	219622
2003	205863	2013	200898
2004	34875	2014	123737
2005	202505	2015	57197
2006	70713	2016	93188
2007	155078	2017	120653
2008	125833	2018	203819
2009	126512	2019	203742

Analyzing Number of Customers:

By analyzing of customer we can predict the number of customers in the future years. Accordingly, we can distinguish the present and the past data by viewing in the visualize output.

Source code:

```python
#define the tools both the pandas and matplotlib
import pandas
import numpy
import matplotlib.pyplot as plt

#reading the csv file with index value=o
table= pandas.read_csv('Stores-data-set3.csv', index_col=0)

df=table.head(20)
print(table)
plt.plot(table)
plt.axis([2000,2019, 1,300000])
#print(table)
plt.legend()
plt.grid()
plt.xlabel('Number of years')
plt.ylabel('Number of Customers')
plt.title('Store DataSets')
plt.show()
```

Importing the **pandas** and the **matplotlib** library to the project file. Creating the table and reading the CSV file in the number of customers has been store. Finally, Plot the table with the X-axis between 2000 to 2019 and giving the name of the a-axis as Number of years and Y-axis as number of Customers.

Analyzing the Customers with percentage:

In this part, I have worked more detail on the plotting the bar graph with describing the each countries sale with percentage and comparing with different percentages in different years of a country. Results are given in Data Analysis Number of Customer per year Table 1 and Data Visualization Number of Customer per year Figure 1.

Table 2:

Country Name	Country Code		2010	2011	2012	2013	2014	2015	2016	2017	2018
Afganistan	AFG		20.6	20.9	19.7	21.1	20.8	19.3	20.3	16	15
Angola	AGO		10.8	10.7	10.7	10.6	10.5	10.3	10.1	10.9	11
Albania	ALB		25.7999	27	28.3	28.7	28.6	28.9	28.4	27	29
Arab World	ARB		25.0022	28.111	29.11132	29.35511	29.331	29	28.1113	29.663	28.565
United Arab Emirates	UAE		10.3	10.63	10.6	10.12	10.98	10.52	10.62	10.79	10.43
Inida	IND		40.32	40.63	40.21	40.52	40.32	40.98	40.87	40.9536	45.595
USA	USA		35.62	35.64	35.12	35.98	36	36.98	36.51	37	37.15
Europe	ERP		60.12	60.32	60.65	60.14	60.52	60.78	60.15	60.95	60.9
Canada	CND		30.21	30.6	30.2	30.5	30.9	30.1	30.8	30.7	30.9

Source code

```
#reading the csv file
country=pd.read_csv('CountrySales.csv', index_col=0)
#reading only first five file
df= country.head(5)
#setting the index value with country code
df=df.set_index(["Country Code"])
dc=dc.set_index(["Country Code"])
#reading only last five file
dc=country.tail(5)

print(country)
#setting the index value to diferrent years
sd=df.reindex(columns=['2010','2011'])
sc=dc.reindex(columns=['2015','2018'],)
#printing the value for example
print(sd)
#plotting the bar graph
de=sc.diff(axis=1)
db=sd.diff(axis=1)
db.plot(kind='bar')
de.plot(kind='bar')
plt.show()
```

After importing **pandas** and **matplotlib** into the project file into the file. Reading the table (2) which is given above and setting the each country code to it for differentiating the customer percentage. Considering the columns of two different years 2010 and 2011 or 2015 and 2018. Columns of 2011 and 2018 shows the initial 5 countries customers, columns of 2015 and 2018 shows the last 5 countries customers. And this differentiation is plotting in bar graph as you can seen in the figure (2)

Analyzing the Sales of a Product:

In this part, I have analyzed the total percentage of a product in a company using Pie-Chart. We are analyzing different sector of a product like Quality, Quantity, Feedback, Sold-out and Not sold.

Table 3:

Sector	Percentage
Quantity	50
Quality	30
feedback	10
sold_out	5
not_sold	5

Source code

```python
piechart= pd.read_csv('customerpie.csv')

print(piechart)
colors = ['darkorange', 'sandybrown', 'darksalmon',
'orangered','chocolate']


plt.pie(piechart['Percentage'],
        labels=piechart['Sector'],
        colors=colors,
        autopct='%.2f')

plt.axis('equal')

plt.show()
```

After importing of pandas and Matplotlib libraries into the project file, I have read the file .CSV file. And I have analyzed the columns and rows which have to visualize. For this result is Pie-char analysis and visualization table (1) and figure (1).

Analyzing Parameters:

Parameters analysis is the process of weighing risk, expressed as variance, against expected return. Investors use Parameters analysis to make decisions about which financial instruments to invest in, based on how much risk they are willing to take on in exchange for different levels of reward. Parameters analysis allows investors to find the biggest reward at a given level of risk or the least risk at a given level of return.

Table 4:

Year	customers	Country
2000	151315	Afghanistan
2001	202307	Angola
2002	37392	Albania
2003	205863	Arab World
2004	34875	United Arab Emirates
2005	202505	Inida
2006	70713	USA
2007	155078	Europe
2008	125833	Canada

Source code

```python
df = pd.read_csv ('stores-data-set3-countries.csv', index_col=0)

# block 1 - simple stats
mean1 = df['customers'].mean()
sum1 = df['customers'].sum()
max1 = df['customers'].max()
min1 = df['customers'].min()
count1 = df['customers'].count()
median1 = df['customers'].median()
std1 = df['customers'].std()
var1 = df['customers'].var()

# block 2 - group by
groupby_sum1 = df.groupby(['Country']).sum()
groupby_count1 = df.groupby(['Country']).count()




# print block 1
print ('Mean customers: ' + str(mean1))
print ('Sum of customers: ' + str(sum1))
print ('Max customers: ' + str(max1))
print ('Min customers: ' + str(min1))
print ('Count of customers: ' + str(count1))
print ('Median customers: ' + str(median1))
```

```
print ('Standard Deviation of customers: ' + str(std1))
print ('Variance of customers: ' + str(var1))

# print block 2
print ('Sum of values, grouped by the Country: ' + str(groupby_sum1))
```

In this part, I have Analyze the data Mean, Summation, Maximum, Minimum, Count, Median, Standard deviation, Variance using pandas. Results are given in Parameter Analysis Table 1:.

OUTPUTS

Data Analysis Number of Customer per year Table 1:

Year	Customers
2000	151315
2001	202307
2002	37392
2003	205863
2004	34875
2005	202505
2006	70713
2007	155078
2008	125833
2009	126512
2010	207499
2011	112238
2012	219622
2013	200898
2014	123737
2015	57197
2016	93188
2017	120653
2018	203819
2019	203742

As you can see the chart lowest numbers of the customers are in the year 2004 which is 34875 and the highest numbers of customers are in the year 2012 which is 219622.

**Data Visualization Number of Customer per year Figure 1:**

Store DataSets
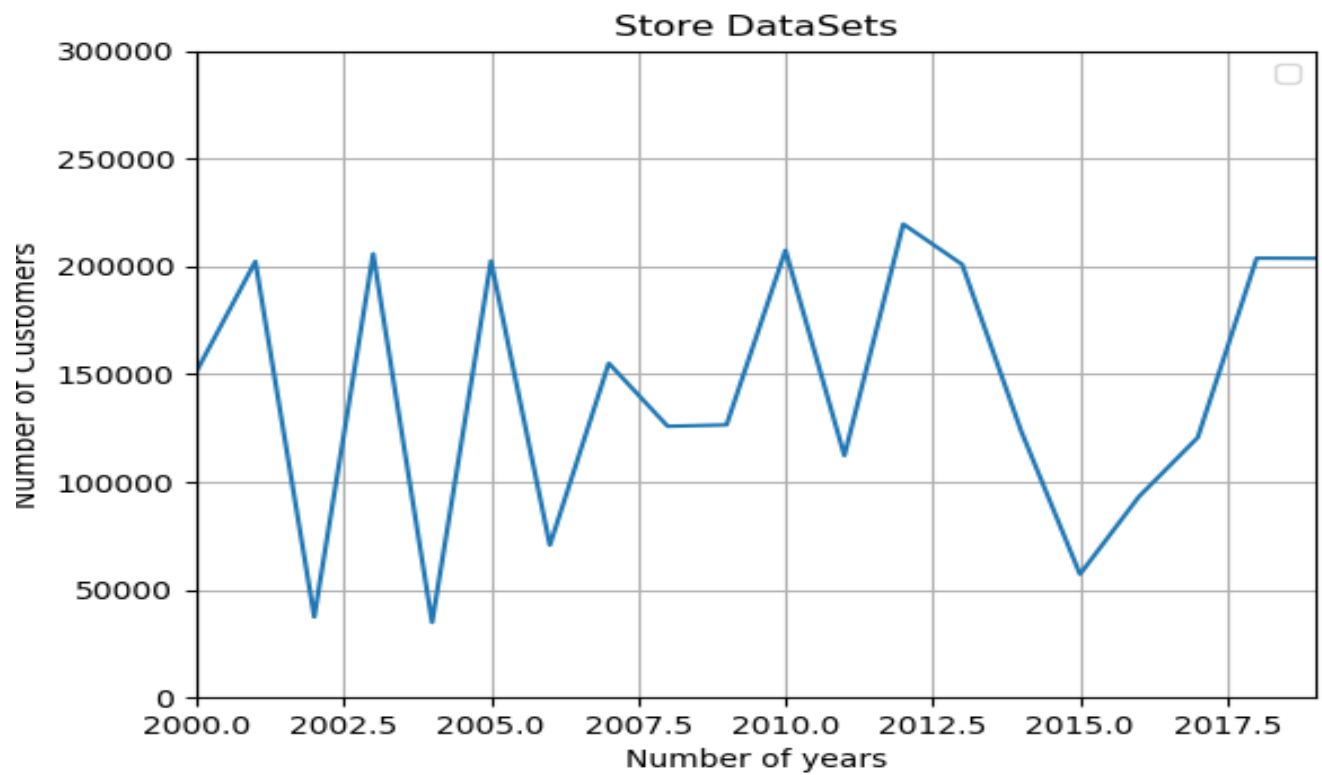
Data Analysis of Percentage Table 1:

Country Code	2010	2011
AFG	20.6000	20.900
AGO	10.8000	10.700
ALB	25.7999	27.000
ARB	25.0022	28.111
UAE	10.3000	10.630

Data Analysis of Percentage Table 2:

Country Code	2015	2018
UAE	10.52	10.430
IND	40.98	45.595
USA	36.98	37.150
ERP	60.78	60.900
CND	30.10	30.900

Data Visualization of percentage Figure 1:
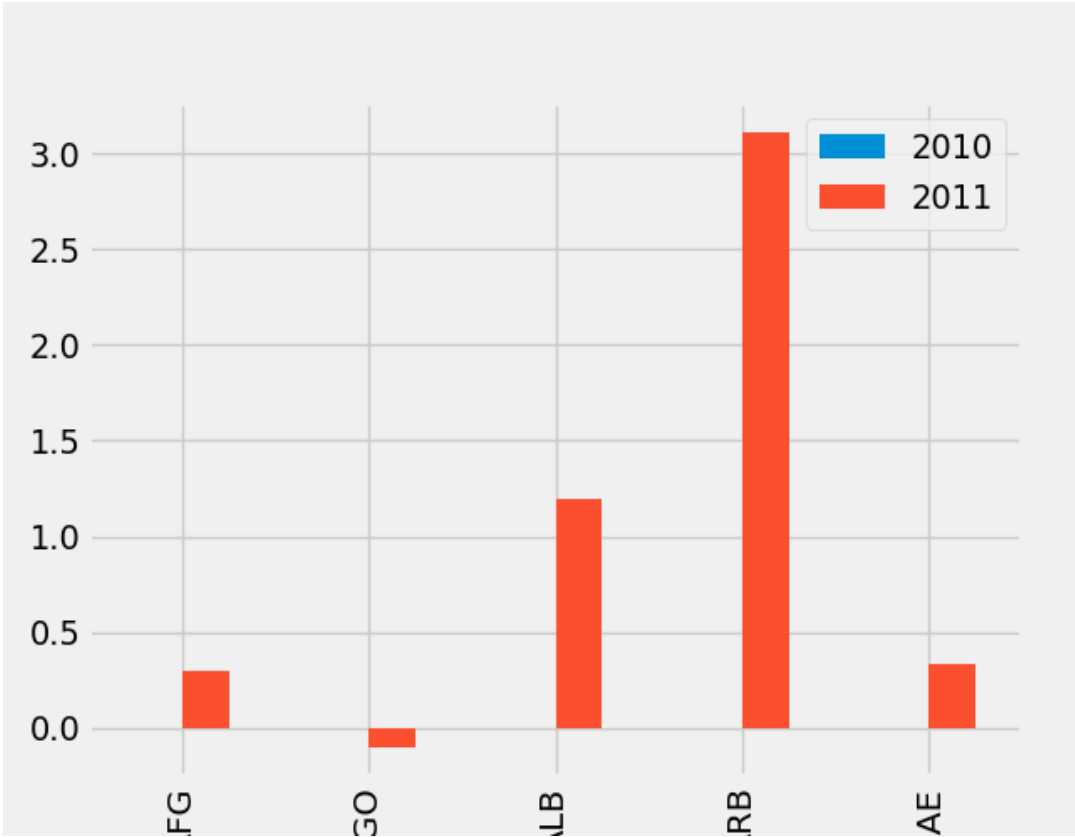
Data Visualization of percentage Figure 2:
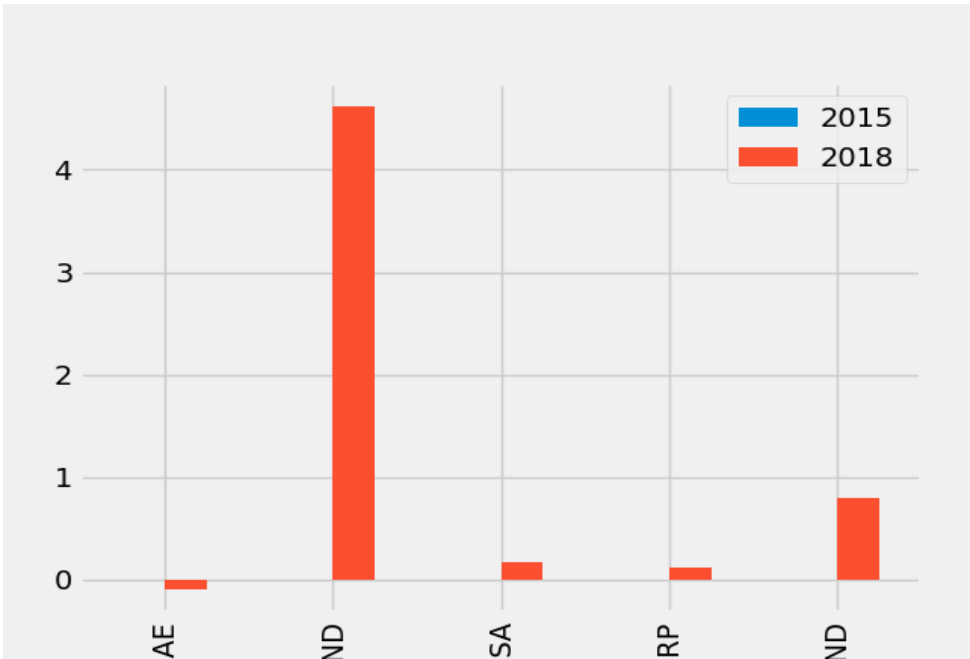
Data Analysis of Pie-chart table 1:

	Sector	Percentage
0	Quantity	50
1	Quality	30
2	feedback	10
3	sold_out	5
4	not_sold	5

Data Visualization of Pie-chart figure1:
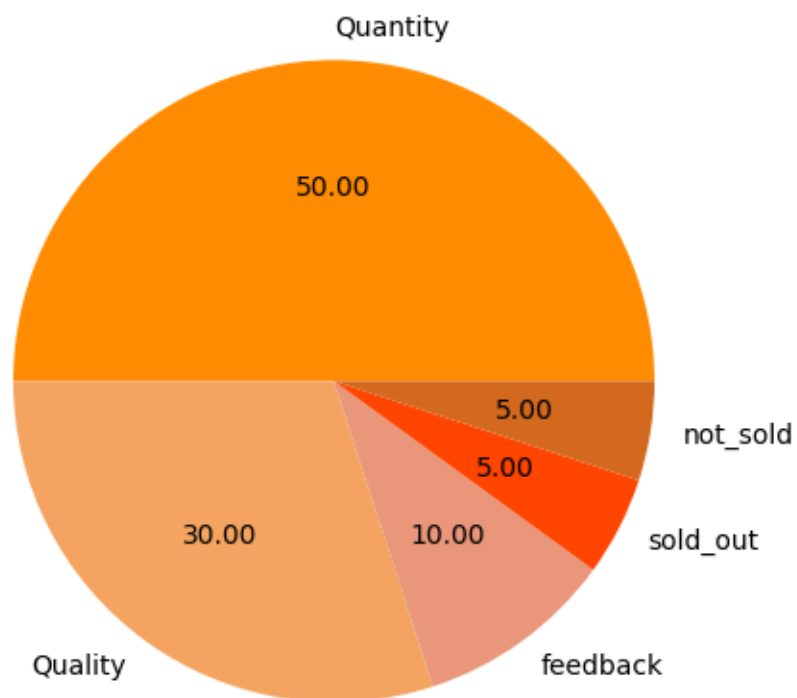
Parameter Analysis Table 1:

Mean customers	131764.55555555556
Sum of customers	1185881
Max customers	205863
Min customers	37392
Count of customers	9
Median customers	151315.0
Standard Deviation	69329.83123466966
Variance	4806625499.027778

Sum of values, grouped by the Country:

Country	customers
Afganistan	151315
Albania	37392
Angola	202307
Arab World	205863
Canada	125833
Europe	155078
India	202505
USA	70713
United Arab Emirates	34875

Conclusion

In this project, we have analyzed and visualized the data of number of customers, the Customers with percentage using Bar Graph, the Sales of a Product using Pie-chart and Analyzing parameters.