Full Name: Daler Khakimov

**Installing a Jupyter Notebook.**

<https://docs.jupyter.org/en/latest/install.html>

**Downloading data.**

<https://github.com/MJey16/dataSource/blob/main/tech-purchase.csv>

**Description.**

A dataset containing tech purchase details

Tasks

1. Cleaning.

*Questions:*

* *How many rows have null values (if there is a null value, that’s also counted)? 545*
* *How many orders have been made (not including nulls)? 355*

*Tasks:*

1. *Show all of the null containing rows*

Code:

df.loc[df["Product"].isna()]

df.loc[df["Product"] == "Product"]

1. *Drop rows that contain any null values*

Code:

df = df.loc[df["Product"].notna()]

df = df.loc[df["Product"] != "Product"]

You can try to consume the data using Jupyter Notebook

1. Restructuring the dataset.

*Tasks:*

1. *Using our (cleaned) dataset, create a new column that would contain only the months from the* ***Order Date*** *column.*

Code:

df["Order Date"] = df["Order Date"].astype("str")

df["Month"] = df["Order Date"].str[0:2]

1. *Convert the month numbers into characters (f.e: 1 into January, 2 into February, and so on).*

Code:

import numpy as np

df.Month = pd.to\_numeric(df.Month, errors='coerce').fillna(0).astype(np.int64)

month = {1:"Jan",2:"Feb",3:"Mar",4:"Apr",5:"May",6:"Jun",7:"Jul",8:"Aug",9:"Sep",10:"Oct",11:"Nov",12:"Dec"}

list = []

def reMonth(mon):

return month[mon]

for index, row in df.iterrows():

list.append(reMonth(row["Month"]))

df.Month = list

1. *Create a new column that would contain only the state names taken from the* ***Purchase Address*** *column.*

Code:

df["Address"] = df["Purchase Address"].astype(str)

df["Address"] = df["Purchase Address"].str[-8:-6]

1. *Convert the* ***Quantity Ordered*** *column’s values into an integer type and the* ***Price Each*** *column’s values into a floating type.*

*Done, while resaving the df*

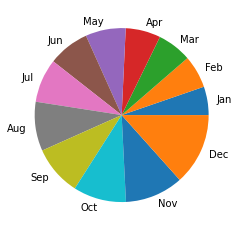
1. Get the useful information.

*Questions:*

1. *Which month has the highest number of sales (in amounts)? And how much was that?*

Dec = 28114 Dec = 4613443.34

1. *Plot a bar plot that would represent the sales info for each month.*

**

1. *Which state has ordered the least and which the highest?*

*Low:*

|  |  |  |  |
| --- | --- | --- | --- |
| **ME** | 2750 | 447189.25 | 449758.27 |
|  |  |  |  |

*High:*

| **Quantity Ordered** | **Price Each** | **Overall** |
| --- | --- | --- |
| **Address** |  |  |  |
| **CA** | 83528 | 13632896.97 | 13714774.71 |

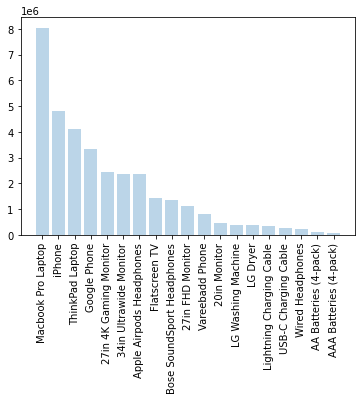
1. *At what time of the day the highest amount of money is spent on sales?*

*At evening (the highest 19:01 -* 54503.14

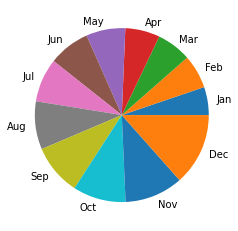
1. Time to plot.

*Tasks:*

1. *Plot the bar chart that represents different items bought on the x-axis and how many of them were sold on the y-axis.*

**

1. *Plot a pie char that represents each month as the piece of the pie, showing how much sales are generally done.*



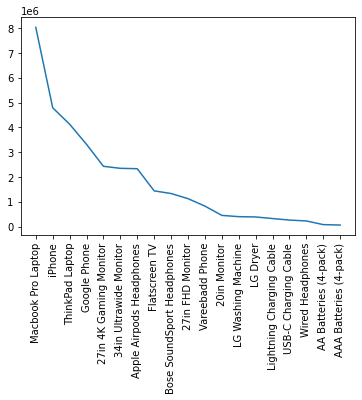
1. More info with the plots.

*Tasks:*

1. *Prove the answer to the question four in the third part (Get the useful information).*

*Sum of evening top 5 =* 268894.41

1. *Referring to the question 1 of the 4th part (Time to plot), plot that bar chart again but with adding a line chart inside onto it representing the price of each item.*



Score: \_\_/5

You can try to consume the data using Jupyter Notebook