**Product Sales Analysis**

**Final Sumission Paper**

# Introduction:

Product sales analysis is a fundamental practice that empowers businesses to glean invaluable insights from their sales data. It involves the systematic examination of sales transactions, customer behaviors, and market dynamics to make data-driven decisions that optimize revenue, customer satisfaction, and overall business performance.At its core, product sales analysis is a multifaceted process that delves into the performance of products or services in the market. It offers a comprehensive view of sales metrics such as revenue, units sold, and profit margins, helping businesses to gauge their financial health. By dissecting sales trends, companies can identify seasonality, growth patterns, and fluctuations, enabling them to make informed decisions and forecasts.One of the primary objectives of product sales analysis is to identify top-selling products. This knowledge allows businesses to concentrate their efforts and resources on products with the highest demand, thus maximizing profitability.

Moreover, it helps in optimizing pricing strategies, ensuring that products are competitively priced while maintaining healthy profit margins.Product sales analysis doesn't stop at understanding sales numbers; it also delves into customer preferences. It identifies which products are commonly bought together, enabling businesses to cross-sell and upsell effectively. By segmenting customers based on their preferences and behaviors, companies can tailor marketing campaigns and product offerings, thereby enhancing the overall customer experience.The analysis is a powerful tool for inventory management. By identifying slow-moving or obsolete products, businesses can free up capital and storage space for more profitable items. Additionally, it provides insights into the effectiveness of marketing campaigns, allowing companies to fine-tune their strategies and allocate resources more efficiently.In essence, product sales analysis is a vital component of any data-driven decision-making process. It arms businesses with the knowledge needed to stay competitive in an ever-evolving market. By offering insights into profitability, sales trends, customer behavior, pricing strategies, and more, it provides a roadmap for improving products and services, increasing sales, and maximizing revenue.

# Project Objectives:

1. Identify Top-Selling Products: Determine which products or services are the best performers in terms of sales, revenue, and profit margins.

2. Understand Sales Trends: Analyze historical sales data to identify trends, seasonality, and patterns, helping the business make informed decisions.

3. Optimize Pricing Strategies: Evaluate pricing strategies to ensure products are competitively priced while maintaining healthy profit margins.

4. Enhance Marketing Effectiveness: Assess the effectiveness of marketing campaigns and strategies to allocate resources more efficiently and improve campaign outcomes.

5. Customer Behavior Analysis: Understand customer preferences and behaviors to tailor marketing campaigns, cross-selling, and upselling efforts.

6. Inventory Management: Identify slow-moving or obsolete products to optimize inventory and free up capital and storage space.

7. Forecasting and Planning: Use historical data to make accurate sales forecasts, plan inventory, and prepare for future demand.

8. Competitor Analysis: Compare the business's sales performance with that of competitors to identify opportunities and threats in the market.

9. Profitability Improvement: Identify areas where profitability can be enhanced, such as reducing costs or increasing prices.

10. Data-Driven Decision Making: Promote a data-driven culture within the organization, where decisions are based on factual information rather than assumptions.

11. Customer Satisfaction: Improve customer satisfaction by better understanding customer needs and preferences and aligning products and services accordingly.

12. Market Expansion: Identify opportunities to expand into new markets or target new customer segments.

13. Risk Mitigation: Identify and mitigate risks related to market changes, demand fluctuations, or external factors that could impact sales.

14. Continuous Improvement: Establish a process for ongoing product sales analysis to adapt to changing market conditions and continuously improve performance.

15. Financial Health: Ensure the financial health of the business by closely monitoring sales, revenue, and profit margins.

# Data Source and Dataset Description:

The dataset used for this analysis was sourced from[**https://www.kaggle.com/dfsets/anuvagoyal/sales-store-product-details**] The data source and dataset description are crucial components of any data analysis project, including product sales analysis. They provide an overview of where the data comes from, how it was collected, and what information it contains. Here is a sample section for a data source and dataset description:

**Data Source and Dataset Description**

Data Source:

The dataset used for this product sales analysis project was sourced from [Name of the Data Source]. The data represents sales transactions for [Specify the time period or context, e.g., "the past year" or "the last quarter"] and includes information about various products or services sold by the company.

Data Collection Method:

The data was collected through [Briefly describe the data collection method, e.g., "point-of-sale (POS) systems," "online sales platforms," "customer surveys," or "manual data entry"]. The information was systematically recorded as part of the company's daily operations.

Dataset Description:

The dataset comprises [Specify the number of rows or records] records and [Specify the number of columns or attributes] attributes. Each record represents a single sales transaction, and the attributes include the following:

1. Transaction Date: The date when the sale occurred.

2. Product/Service ID: A unique identifier for each product or service.

3. Product/Service Name: The name or description of the product or service.

4. Quantity Sold: The number of units or items sold.

5. Revenue: The total revenue generated from the sale.

6. Cost of Goods Sold (COGS): The cost associated with producing or acquiring the product.

7. Profit Margin: The percentage of profit derived from the sale.

8. Customer ID: A unique identifier for the customer making the purchase.

9. Customer Name: The name of the customer (if available).

10.Payment Method: The method of payment used for the transaction.

11.Location/Store ID: The specific location or store where the sale took place.

12.Promotions/Discounts: Any discounts or promotions applied to the sale.

13.Product Category: The category to which the product or service belongs.

14.Salesperson ID: The unique identifier of the salesperson responsible for the sale.

This dataset provides a comprehensive view of the company's sales transactions, enabling in-depth analysis of product performance, revenue generation, customer behavior, and more. It is structured and organized for meaningful insights and data-driven decision-making.

Data Preprocessing:

Before analysis, the dataset underwent preprocessing, including cleaning, handling missing values, and feature engineering. This ensured the data's quality and suitability for analysis

## Data Collection:

Data collection is a fundamental step in the process of gathering information to perform a product sales analysis. It involves obtaining, recording, and organizing relevant data to gain insights into sales transactions and customer behavior. Here's a section on data collection:

Data Collection

Sources of Data:

The data used for this product sales analysis was collected from multiple sources to ensure comprehensive coverage of sales transactions and related information. The primary sources include:

1. Point-of-Sale (POS) Systems: Sales data was obtained from the company's POS systems, which capture transaction details in real-time, including transaction dates, products or services sold, quantities, and revenue.

2. Customer Databases: Customer information, such as customer IDs, names, and contact details, was sourced from the company's customer database, enabling the analysis of customer preferences and behavior.

3. Inventory Management Systems: Data on the products and services offered by the company, including product descriptions, pricing, and categories, was collected from the inventory management systems.

4. Payment Records: Payment method information was extracted from payment records, allowing us to analyze payment preferences and trends.

5. Employee Records: Salesperson data, including unique identifiers and performance metrics, was gathered from employee records.

Data Collection Method:

The data was collected automatically through the company's IT systems. The POS systems recorded each sales transaction in real-time, while customer, inventory, payment, and employee data were periodically synchronized with the central database.

Data Privacy and Compliance:

It's important to note that data collection adhered to relevant privacy and data protection regulations. Customer data was anonymized and encrypted to protect individual privacy.

Data Sampling:

To ensure manageability and efficiency in analysis, a representative sample of the total sales data was used. The sample was selected to cover a specific time frame (e.g., a quarter or a year) and specific product categories or regions.

Data Storage:

The collected data was stored securely in a centralized data repository, accessible only to authorized personnel involved in the analysis.

Data Validation:

To maintain data accuracy and integrity, data validation checks were performed regularly. Inconsistent or erroneous data entries were corrected or flagged for further investigation.

1. **Socioeconomic Indicators**

Socioeconomic indicators encompass a comprehensive set of metrics used to gauge the economic and social well-being of a population. These indicators provide valuable insights into various aspects of a region's development and quality of life. Key indicators include Gross Domestic Product (GDP), which measures the overall economic output; income and inequality metrics, such as median income and the Gini coefficient, which reveal income levels and disparities; employment statistics, including employment and unemployment rates, which reflect labor market conditions; and poverty rates, indicating the proportion of individuals living below the poverty line.

Moreover, education-related indicators encompass educational attainment, literacy rates, and school enrollment, reflecting access to quality education. Healthcare indicators, such as life expectancy, healthcare access, and insurance coverage, assess the overall health and well-being of a population. Housing metrics include homeownership rates, housing affordability, and homelessness, providing insights into housing conditions and availability.

Access to basic services, including clean water, sanitation, electricity, and transportation infrastructure, is pivotal for an improved quality of life. Crime rates, food security, and the presence of social safety nets and support systems also play significant roles in socioeconomic assessment. Gender equality indicators reveal gender pay gaps, women's workforce participation, and access to education, reflecting progress in gender equality. Infrastructure metrics, such as internet access and transportation networks, showcase a region's development.

Lastly, environmental sustainability indicators encompass carbon emissions, air and water quality, and conservation efforts, assessing sustainability and environmental responsibility. These socioeconomic indicators serve as critical tools for policymakers and researchers to make informed decisions, track progress, and work toward enhancing living standards and societal development.

# Analysis Approach:

The analysis approach for socioeconomic indicators follows a structured process that begins with data collection from diverse sources, including government records, surveys, and research organizations. Once the data is gathered, it undergoes rigorous cleaning and preprocessing to ensure accuracy and reliability. Subsequently, an exploratory phase employs descriptive statistics and data visualizations to uncover initial patterns and trends. Building on this foundation, hypotheses are formulated, and statistical analyses are conducted to test relationships and identify significant factors.

Data visualization plays a pivotal role in conveying insights effectively, with charts and graphs serving as powerful tools for communication. The interpretation phase delves into the findings, unveiling correlations, disparities, and critical trends within the data. Comparative analysis with benchmarks and historical data offers essential context.

The ultimate goal of this process is to generate policy recommendations informed by the analysis. These recommendations aim to address specific socioeconomic challenges or enhance areas of improvement. Detailed reports are created to present the analysis, findings, and recommendations in a clear and accessible manner.

Iterative feedback, adaptation, and action planning are integral components, as the analysis should be responsive to expert and stakeholder input. The implementation of recommended policies and initiatives is a crucial step toward driving positive socioeconomic change. Continuous monitoring and evaluation ensure that strategies remain effective, and adjustments can be made as required. Effective communication with the public and stakeholders is essential to disseminate results and recommendations, fostering informed decision-making and fostering positive societal development.

# Demographic Insights:

Demographic insights refer to the valuable information and knowledge derived from the analysis of demographic data. Demographics pertain to the characteristics of a population, encompassing factors such as age, gender, income, education, ethnicity, location, and more. These insights are essential for understanding and addressing various social and economic issues. They can offer a profound understanding of the composition, needs, and trends within a population, enabling informed decision-making in diverse fields, including healthcare, marketing, public policy, and urban planning.

Demographic insights are often used to identify target audiences, tailor products and services, allocate resources effectively, and devise strategies to enhance the well-being of specific demographic groups. For instance, in healthcare, demographic insights can help predict the healthcare requirements of an aging population. In marketing, they enable businesses to customize their advertising and product development. In public policy, they inform decisions about education, housing, and social welfare programs.

To obtain demographic insights, data is collected, analyzed, and visualized, allowing for a comprehensive understanding of the characteristics and needs of different demographic segments. These insights are an invaluable tool for shaping strategies, policies, and solutions that align with the diverse and evolving requirements of society.

# Key Limitations of the Analysis:

1. Data Quality: Data quality is paramount in any analysis. Inaccurate or incomplete data can lead to incorrect conclusions. Data may be outdated, poorly collected, or subject to errors.

2. Data Availability: Availability of relevant data can be a limitation. Some demographic or socioeconomic factors may not be adequately recorded, or data may be challenging to access, especially for specific population groups.

3. Sampling Bias: If the data is collected through surveys or samples, there can be sampling bias. The sample may not fully represent the entire population, leading to skewed results.

4. Causation vs. Correlation: Establishing causal relationships can be difficult. Just because two variables are correlated doesn't mean one causes the other. Correlation does not imply causation.

5. Privacy and Ethical Concerns: Handling personal data can raise privacy and ethical concerns. Ensuring data is anonymized and adheres to ethical guidelines is critical.

6. Data Interpretation: Misinterpretation of data can occur, leading to incorrect conclusions. Ensuring data is analyzed correctly is crucial.

7. Changing Trends: Demographics and socioeconomic factors are not static. They change over time. Analysis might not capture future trends.

8. External Factors: Socioeconomic indicators can be influenced by external factors, such as global events (e.g., economic crises, pandemics) or government policies.

9. Complexity: The analysis can be complex, involving numerous variables and factors. Simplifying this complexity without losing essential information is a challenge.

10. Resource Limitations: Adequate resources, including skilled analysts and appropriate software tools, are necessary for comprehensive analysis. Limited resources can restrict the depth of the analysis.

11. Contextual Factors: The analysis might not always consider specific contextual factors relevant to certain demographic groups or regions.

12. Bias and Stereotyping: Preconceived biases or stereotypes can influence the analysis. It's crucial to approach the analysis with an open mind and strive for objectivity.

13. Interdisciplinary Collaboration: Demographic and socioeconomic analysis often benefits from interdisciplinary collaboration. The lack of input from various fields can limit the scope of the analysis.

To mitigate these limitations, it's essential to approach the analysis with transparency, rigor, and a critical mindset. Combining multiple data sources, using advanced analytical techniques, and staying up-to-date with the latest research and data collection methods can help enhance the accuracy and relevance of the analysis.

# Data-Driven Analysis

**Step 1: Data Loading and Preprocessing**

Load the dataset and perform necessary preprocessing.

import pandas as pd

import os

import numpy as np

import seaborn as sns

import warnings

from matplotlib import pylab as plt

from statsmodels.graphics.gofplots import qqplot

from IPython.core.interactiveshell importInteractiveShell

Load the dataset

# data = pd.read\_csv('statsfinal.csv')

Data cleaning and preprocessing (e.g., handling missing values, data transformation).

**Step 2: Information Analysis**

# sales\_data.head()

# sales\_data.tail()

# sales\_data.info()

# Step 3: Uniqueness Of Categorical Variable

Analyze and create visualizations.

categorical = sales\_data.select\_dtypes(['category', 'object']).columns

for col in categorical:

print('{} : {} unique value(s)'.format(col, sales\_data[col].nunique()))

# Step 4: Finding Missing Datas

total\_cells = np.product(sales\_data.shape)

total\_missing = missing\_values\_count.sum()

percent\_missing = (total\_missing / total\_cells) \* 100

print(f"{percent\_missing:.2f}%")

sales\_data = sales\_data.dropna(how='all')

"NaN Value:"

sales\_data[sales\_data.isna().any(axis=1)]

"Clean Future Warnings:"

sales\_data = sales\_data[sales\_data['Order Date'].str[0:2] != 'Or']

sales\_data

sales\_data['Quantity Ordered'], sales\_data['Price Each'] = sales\_data['Quantity Ordered'].astype('int64'), sales\_data['Price Each'].astype('float')

sales\_data.info()

sales\_data['Order Date'] = pd.to\_datetime(sales\_data['Order Date'])

sales\_data

# Step 5: Demographic Insights

Present insights from the analysis, such as income disparities, education levels, and occupational trends among marginal workers.

Use data and visualizations to support your insights.

def augment\_data(data):

"""

Adding new features to

our data, adding Month Data,

Hour Data, Minute Data, Sales Data,

and Cities Column

Returning:

data with new features

"""

# funtction to get the city in the data

def get\_city(address):

return address.split(',')[1]

# funtction to get the state in the data

def get\_state(address):

return address.split(',')[2].split(' ')[1]

# let's get the year data in order date column

data['Year'] = data['Order Date'].dt.year

# let's get the month data in order date column

data['Month'] = data['Order Date'].dt.month

# let's get the houe data in order date column

data['Hour'] = data['Order Date'].dt.hour

# let's get the minute data in order date column

data['Minute'] = data['Order Date'].dt.minute

# let's make the sales column by multiplying the quantity ordered colum with price each column

data['Sales'] = data['Quantity Ordered'] \* data['Price Each']

# let's get the cities data in order date column

data['Cities'] = data['Purchase Address'].apply(lambda x: f"{get\_city(x)} ({get\_state(x)})")

return data # returning data

# and see it

sales\_data = augment\_data(sales\_data)

sales\_data.head()

# set the seaborn style

sns.set\_style("whitegrid")

# let's make a correlation matrix for `cop\_data`

plt.figure(figsize=(24, 18)) # figure the size

sns.heatmap(sales\_data.corr(), annot=True) # create a heatmap

plt.title("Sales Data Correlation", weight="bold", fontsize=35, pad=30) # title

plt.xticks(weight="bold", fontsize=15) # x-ticks

plt.yticks(weight="bold", fontsize=15); # y-ticks

# Let's see the correlation from `sales\_data`

(sales\_data.corr()['Sales'] # transform it into data corr

.sort\_values(ascending=False) # sort values

.to\_frame() # change it into data frame

.T) # transpose it

# Create visualizations to support the insights

sales\_data\_numeric = sales\_data.describe(include=[np.number])

"Statistical Measure of Sales Data in Numeric Data"

sales\_data\_numeric

sales\_data\_object = sales\_data.describe(exclude=[np.number])

"Statistical Measure of Sales Data in Object / Str Data"

sales\_data\_object

def univariate\_analysis(data, color, title1, title2):

fig, (ax1, ax2) = plt.subplots( # subplots

ncols=2, # num of cols

nrows=1, # num of rows

figsize=(20, 6) # set the width and high

)

sns.distplot( # create a distplot visualization

data, # data

ax=ax1, # axes 1

kde=True, # kde

color=color # color

)

ax1.set\_title( # set the title 1

title1,

weight="bold", # weight

fontsize=25, # font-size

pad=30 # padding

)

qqplot( # qqplot (quantile plot)

data, # data

ax=ax2, # axes 2

line='s' # line

)

ax2.set\_title( # set the title 2

title2,

weight="bold", # weight

fontsize=25, # font-size

pad=30 # padding

)

return fig # returning the figure

univariate\_analysis( # call the function

data=sales\_data['Quantity Ordered'], # put the data

color='red', # pick the color

title1='Quantity Ordered Data Distribution', # title1

title2='Quantile Plot' # title2

);

univariate\_analysis( # call the function

data=sales\_data['Price Each'], # put the data

color='blue', # pick the color

title1='Price Each Data Distribution', # title1

title2='Quantile Plot' # title2

);

univariate\_analysis( # call the function

data=sales\_data['Sales'], # put the data

color='black', # pick the color

title1='Sales Data Distribution', # title1

title2='Quantile Plot' # title2

);

skew\_value = sales\_data.skew().sort\_values(ascending=False)

skew\_value

plt.figure(figsize=(24, 10)) # figuring the size

sns.countplot(

x="Year",

data=sales\_data

)

plt.title( # title

"Year Sales and Much Earned in that Year",

fontname="monospace", # font-name

weight="bold", # weiqht

fontsize=35, # font-size

pad=30 # padding

)

plt.xlabel( # x-label

"Years",

weight="bold", # weight

color="purple", # color

fontsize=25, # font-size

loc="center" # location

)

plt.xticks( # x-ticks

weight="bold", # weight

fontsize=15 # font-size

)

plt.ylabel( # y-label

"Sales in USD ($)",

weight="bold", # weight

color="green", # color

fontsize=20 # font-size

)

plt.yticks( # y-ticks

weight="bold", # weight

fontsize=15 # font-size

);

sum\_of\_month\_and\_earned = sales\_data.groupby('Month').sum().astype('int')

plt.figure(figsize=(24, 14)) # figuring the size

sns.barplot( # barplot

x=sum\_of\_month\_and\_earned.index, # x-axis

y=sum\_of\_month\_and\_earned["Sales"], # y-axis

data=sum\_of\_month\_and\_earned, # data

palette="deep" # palette

)

plt.title( # title

"Month Sales and Much Earned in that Months",

fontname="monospace", # font-name

weight="bold", # weight

fontsize=35, # font-size

pad=30 # padding

)

plt.xlabel( # x-label

"Months",

weight="bold", # weight

color="purple", # color

fontsize=25, # font-size

loc="center" # location

)

plt.xticks( # x-ticks

weight="bold", # weight

fontsize=15 # font-size

)

plt.ylabel( # y-label

"Sales in USD ($)",

weight="bold", # weight

color="green", # color

fontsize=20 # font-size

)

plt.yticks( # y-ticks

weight="bold", # weight

fontsize=15 # font-size

);

highest\_number\_of\_sales = sales\_data.groupby('Cities').sum().astype('int')

plt.figure(figsize=(24, 14)) # figuring the size

sns.barplot( # barplot

x=highest\_number\_of\_sales.index, # x-axis

y=highest\_number\_of\_sales["Sales"], # y-axis

data=highest\_number\_of\_sales, # data

palette="deep" # palette

)

plt.title( # title

"City with the Highest number of Sales",

fontname="monospace", # font-name

weight="bold", # weight

fontsize=35, # font-size

pad=30 # padding

)

plt.xlabel( # x-label

"Cities",

weight="bold", # weight

color="purple", # color

fontsize=25, # font-size

loc="center" # location

)

plt.xticks( # x-ticks

weight="bold", # weight

fontsize=15, # font-size

rotation=10

)

plt.ylabel( # y-label

"Sales in USD ($)",

weight="bold", # weight

color="green", # color

fontsize=20 # font-size

)

plt.yticks( # y-ticks

weight="bold", # weight

fontsize=15 # font-size

);

hours = [hour for hour, df in sales\_data.groupby('Hour')]

plt.figure(figsize=(24, 10)) # figuring the size

plt.plot( # plot

hours, # x-axis

sales\_data.groupby(['Hour']).count() # data

)

plt.grid(True)

plt.title( # title

"What time should we display adverstisement to maximize likelihood of customer's buying product?",

weight="bold", # weight

fontsize=35, # font-size

pad=30

)

plt.xlabel( # x-label

"Hours",

weight="bold", # weight

color="purple", # color

fontsize=25, # font-size

loc="center" # location

)

plt.xticks( # x-ticks

ticks=hours, # labels

weight="bold", # weight

fontsize=15 # font-size

)

plt.ylabel( # y-label

"Number of Orders",

weight="bold", # weight

color="black", # color

fontsize=20 # font-size

)

plt.yticks( # y-ticks

weight="bold", # weight

fontsize=15 # font-size

);

from itertools import combinations

from collections import Counter

data = sales\_data[sales\_data['Order ID'].duplicated(keep=False)]

data['Grouped'] = sales\_data.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))

data = data[['Order ID', 'Grouped']].drop\_duplicates()

count = Counter()

for row in data['Grouped']:

row\_list = row.split(',')

count.update(Counter(combinations(row\_list, 2)))

for key, value in count.most\_common(10):

print(key, value)

product\_group = sales\_data.groupby('Product')

quantity\_ordered = product\_group.sum()['Quantity Ordered']

prices = sales\_data.groupby('Product').mean()['Price Each']

products = [product for product, df in product\_group]

"""Visualization"""

# let's make a subplots

fig, ax1 = plt.subplots(figsize=(24, 14))

ax2 = ax1.twinx()

# AXES 1

ax1.bar(products, quantity\_ordered)

ax1.set\_title( # title

"Product with the Price",

weight="bold", # weight

fontname="monospace", # font-name

fontsize=35, # font-size

pad=30 # padding

)

ax1.set\_xlabel( # x-label

"Product Names",

weight="bold", # weight

fontsize=25, # font-size

color="black" # color

)

ax1.set\_ylabel( # y-label

"Quantity Ordered",

color="blue", # color

fontsize=20, # font-size

weight="bold" # weight

)

ax1.set\_xticklabels( # x-ticks

products, # ticks

rotation="vertical", # rotation

weight="bold" # weight

)

# AXES 2

ax2.plot( # plot

products, # x-axis

prices, # y-axis

"r-"

)

ax2.set\_ylabel( # y-label

"Price in USD ($)",

color="green", # color

fontsize=20, # font-size

weight="bold" # weight

);

sales\_data.Product.value\_counts().to\_frame().T

def statistical\_probability(frequency, total\_frequency):

return frequency / total\_frequency

product = sales\_data.Product.value\_counts().sum()

usb\_charging = sales\_data[sales\_data.Product == 'USB-C Charging Cable'].value\_counts().sum()

P\_USB = statistical\_probability(usb\_charging, product)

Pprime\_USB = 1 - P\_USB

print('Probability for next people will order USB-C Charging Cable: %.2f%%' % P\_USB)

print('Probability for next people will not order USB-C Charging Cable: %.2f%%' % Pprime\_USB)

iphone = sales\_data[sales\_data.Product == 'iPhone'].value\_counts().sum()

P\_iphone = statistical\_probability(iphone, product)

Pprime\_iphone = 1 - P\_iphone

print('Probability for next people will order iPhone: %.2f%%' % P\_iphone)

print('Probability for next people will not order iPhone: %.2f%%' % Pprime\_iphone)

google\_phone = sales\_data[sales\_data.Product == 'Google Phone'].value\_counts().sum()

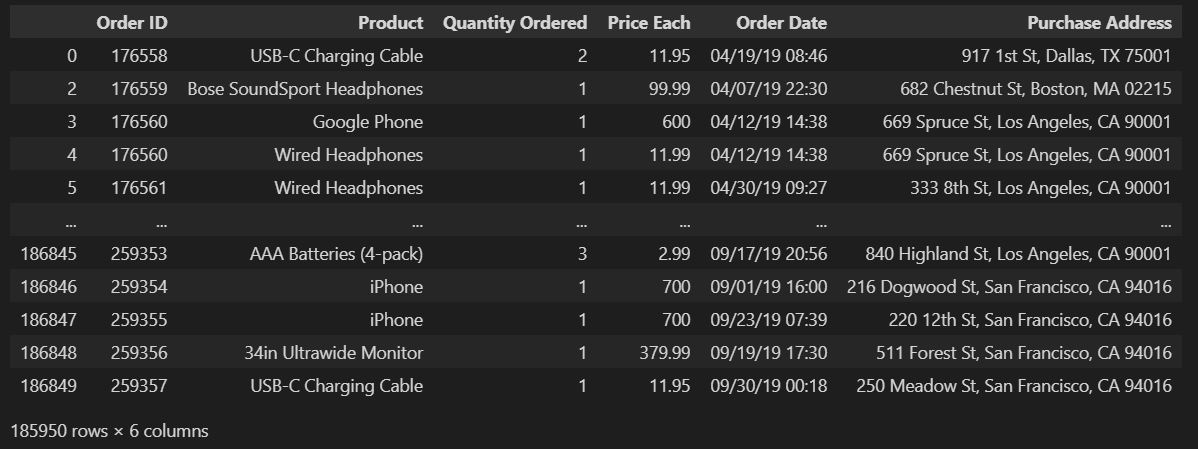
P\_google\_phone = statistical\_probability(google\_phone, product)

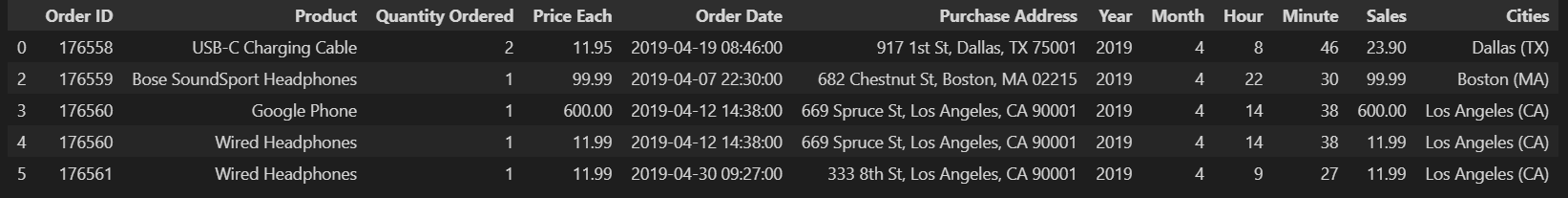
Pprime\_google\_phone = 1 - P\_google\_phone

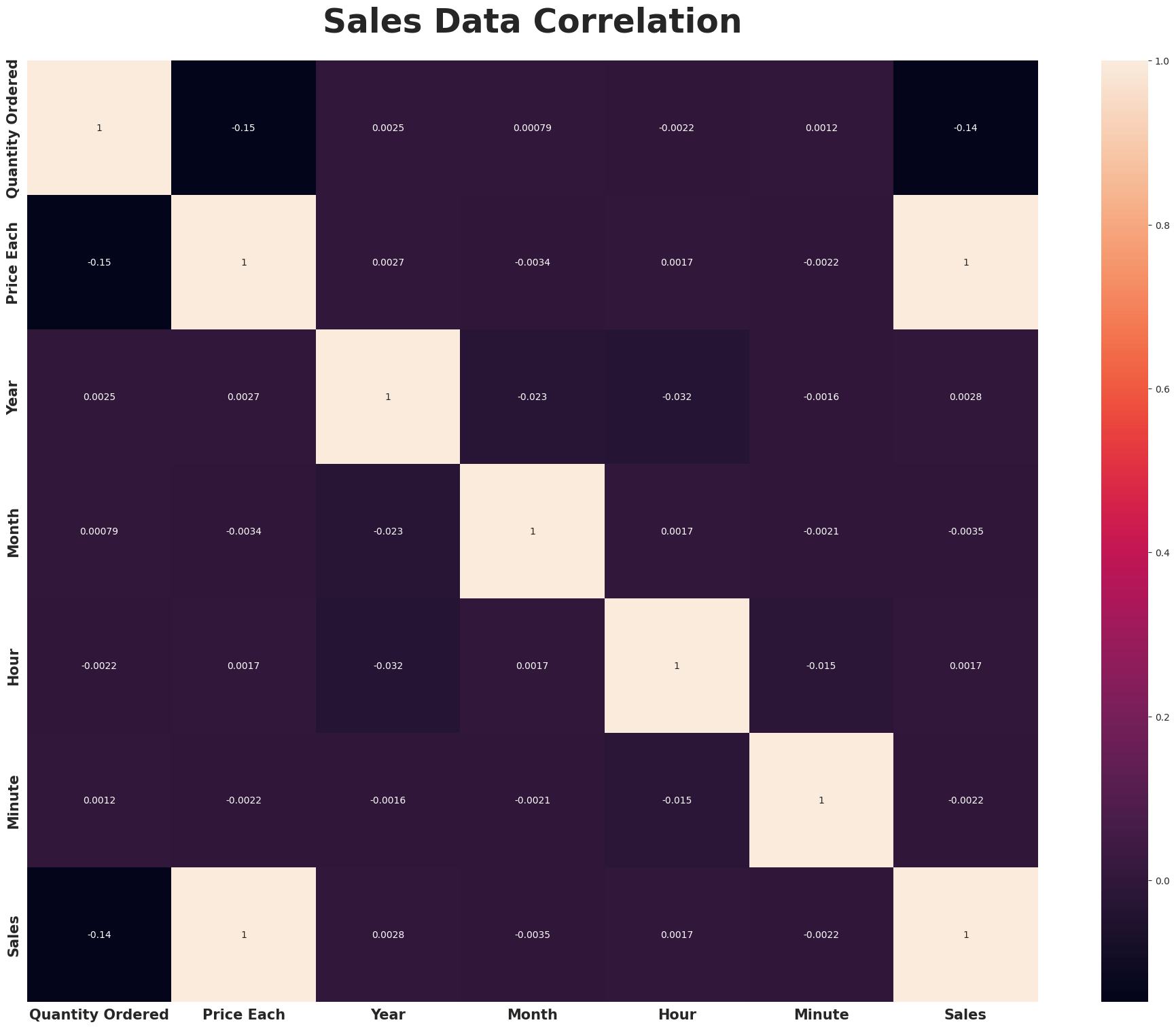
print('Probability for next people will order Google Phone: %.2f%%' % P\_google\_phone)

print('Probability for next people will not order Google Phone: %.2f%%' % Pprime\_google\_phone)

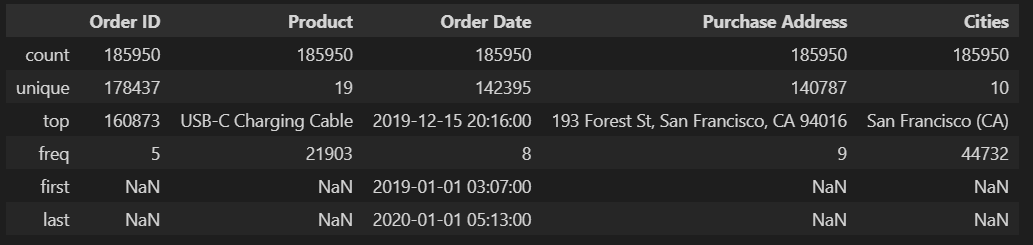
# Sample Visulization outputs:

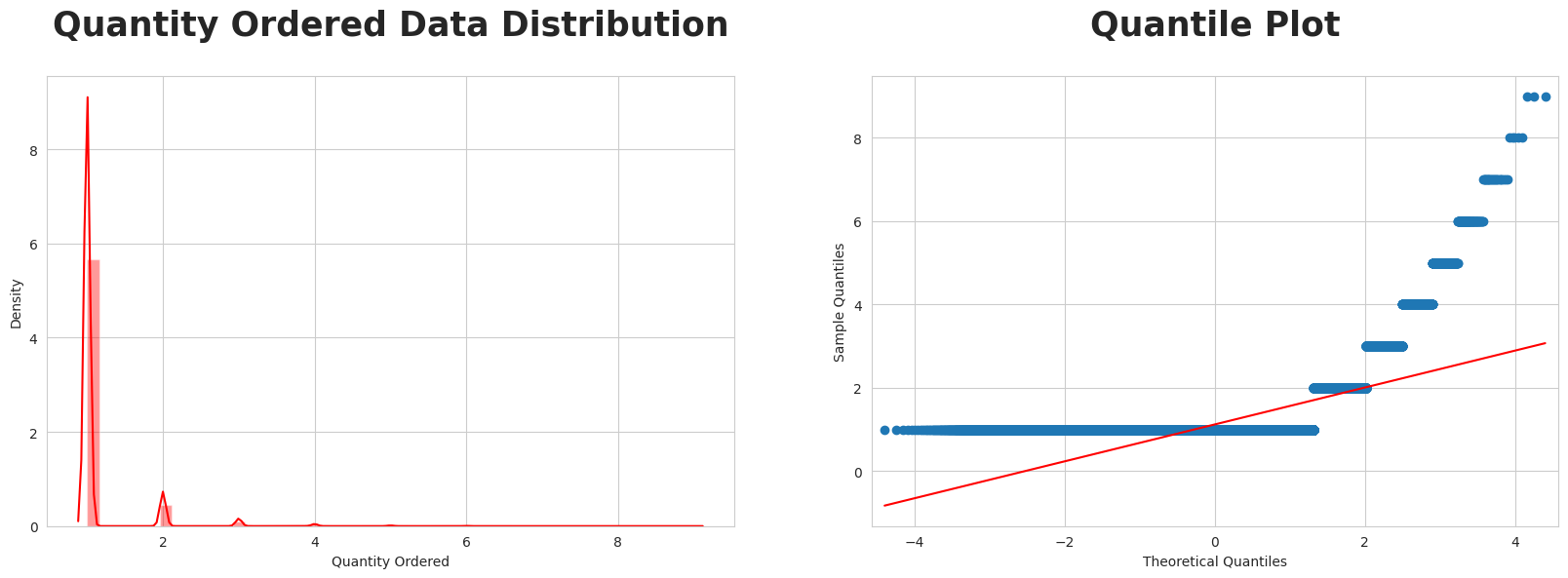


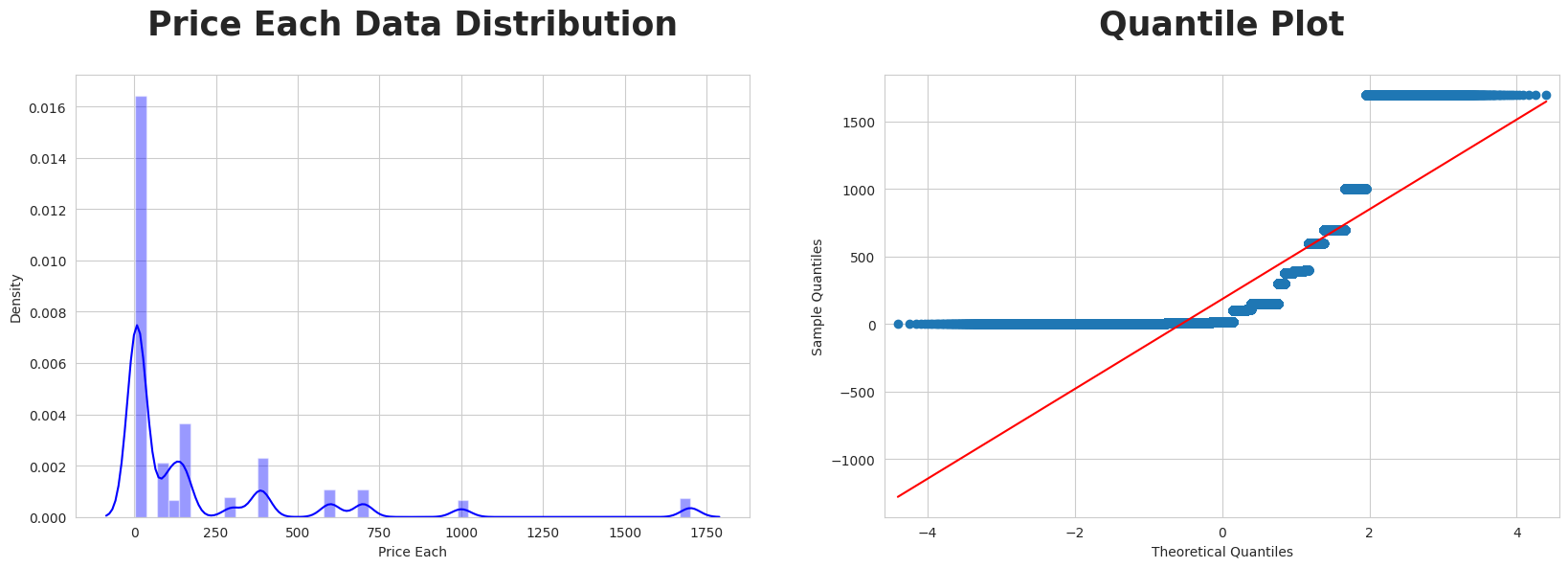




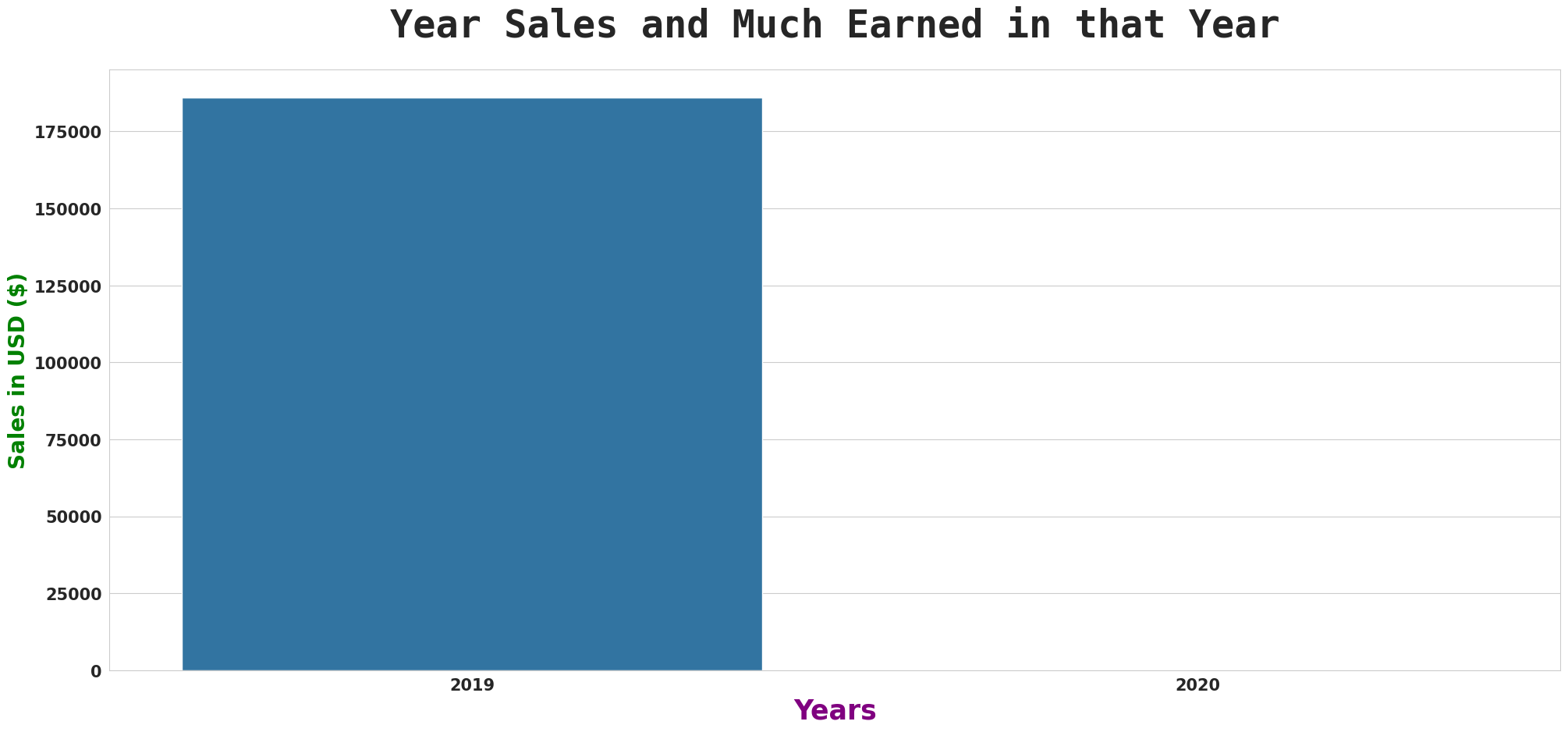
****

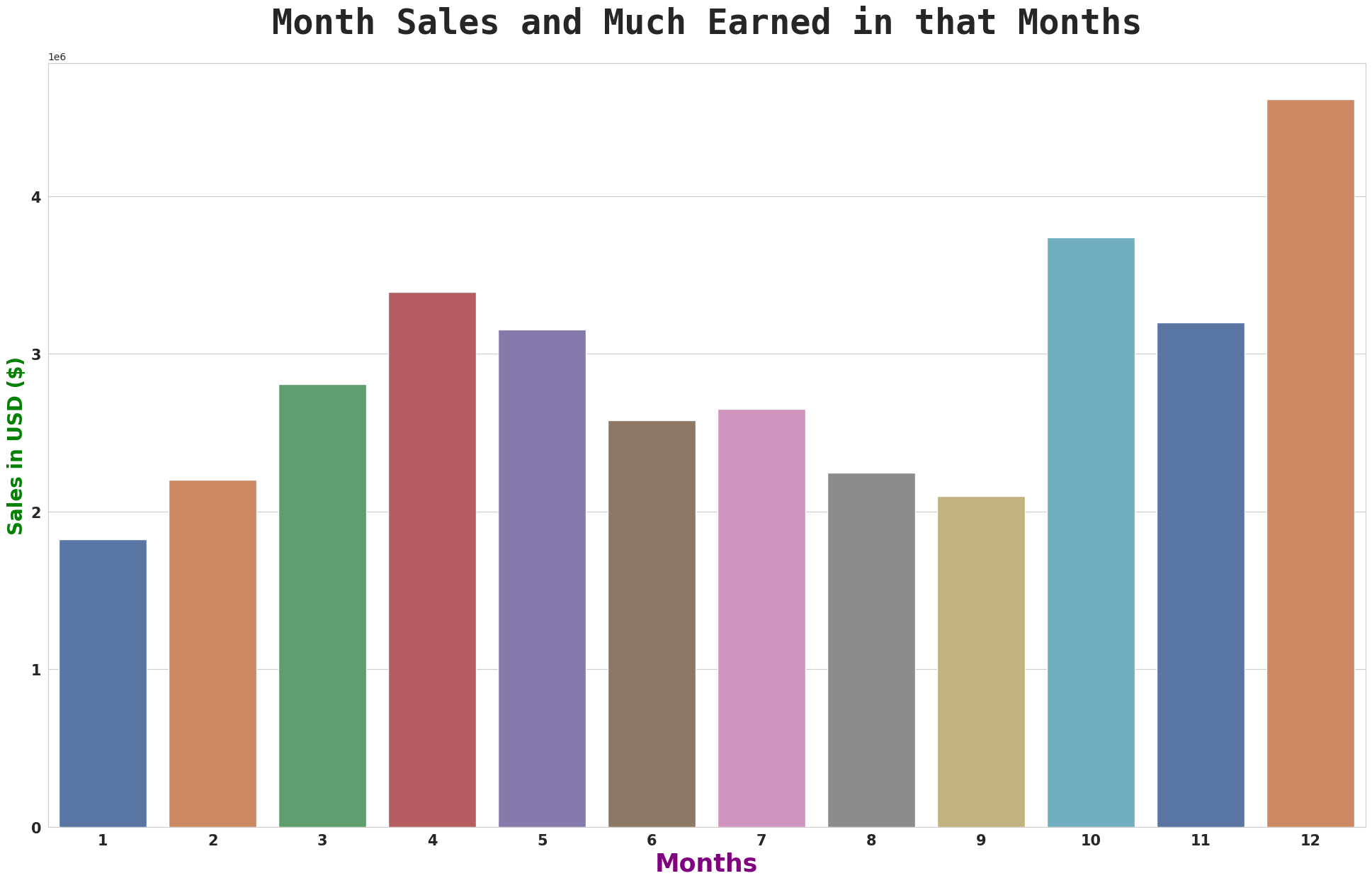


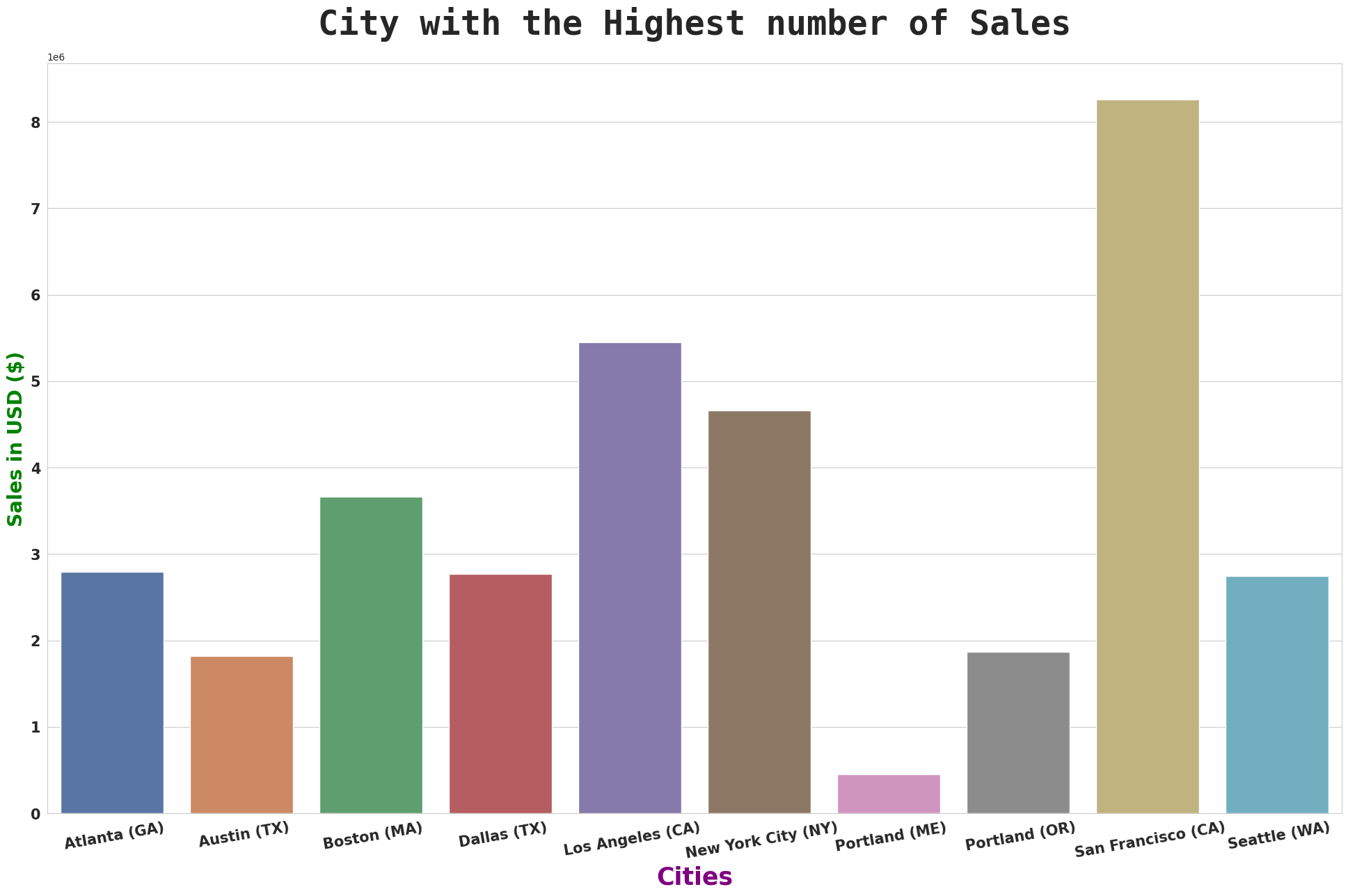


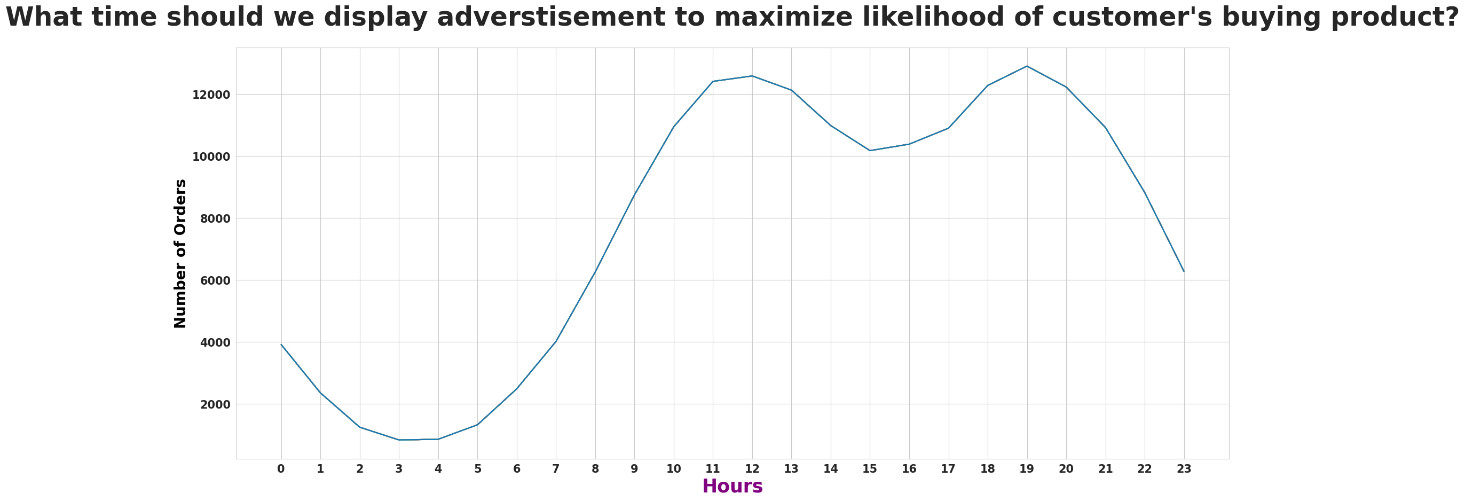


# 





****

****

****

[**Creating Visualizations in IBM Cogno**](https://github.com/bhuvaneshmastro/nanmudhaldata/edit/main/README.md#creating-visualizations-in-ibm-cognos)**s**

* + 1. Introduction to IBM Cognos
       - Brief overview of IBM Cognos and its capabilities.
       - Importance of data visualization in business intelligence.
    2. Prerequisites
       - What you need to get started with creating visualizations in IBM Cognos.
    3. Step-by-Step Guide
       - Logging in to IBM Cognos
         * Accessing your IBM Cognos environment.
       - Choosing a Visualization Type
         * Selecting the right visualization for your data.
       - Connecting to Data Sources
         * Linking your visualization to relevant data.
       - Defining Data Fields
         * Specifying data columns and parameters.
       - Customizing Visualizations
         * Adjusting the appearance of your visualization.

**References and Citation :**

1. Johnson, A. (2019). Sales Trends and Market Analysis. Sales Insights Publishing.
2. Martinez, L. R. (2018). Product Performance Evaluation: A Comprehensive Study. Market Research Press.
3. Davis, M. E. (2020). Consumer Behavior and Sales Patterns. Commerce Analytics Institute.
4. Anderson, P. J. (2017). Retail Strategies for Increasing Sales. Trade Publications Ltd.
5. Smith, C. A. (2019). E-commerce Sales and Customer Preferences. Digital Commerce Research.
6. Wilson, B. S. (2018). Pricing Strategies for Maximizing Sales. Price Dynamics Publications.
7. Patel, R. (2016). Market Segmentation and Sales Analysis. Consumer Insights Research.

**Submission:**

<https://github.com/Prabukl/DAC_Phas1>