Team 3: Hafeezah, Connor, Babatola, Selda, Iqra

Chocolate Sales Analysis

Built for business intelligence and decision-making, this tool enables users to uncover sales trends, understand factors affecting chocolate sales and optimize business strategies for increasing revenue.

Overview

Providing an in-depth analysis of dataset and uncovering key insights into which demographics are at risk of having thyroid cancer:

- **Hypothesis 1**: Age and gender influence thyroid cancer risk, with older individuals and females having a higher probability
  - **Validation**: Use box plots and regression analysis to explore how cancer risk varies across different age groups and gender distributions

- **Hypothesis 2**: Certain countries and ethnicities have a higher prevalence of thyroid cancer due to genetic and environmental factors
  - **Validation**: Conduct geospatial analysis and visualize the distribution of thyroid cancer cases across different regions.

Planning & Design

Ideation

Project Goal: Build interactive dashboards for data analysis

Business Requirement: understand factors affecting chocolate sales and optimize business strategies for increasing revenue, improving sales efficiency, and enhancing customer targeting.

Target Audience: Marketing professionals, Sales Managers, Product Managers, Retail Partners, Data Analysts, Executives and Business Owners.

Design

User Stories: "As a Business Analyst / Sales Manager, view monthly sales trends to identify seasonal demand fluctuations and compare sales performance across different countries

Intuitive UI: Clean layouts, easy navigation

Accessibility: Readable colours, labeling

Interactivity: Clickable filters, zoomable charts and maps

Hypothesis: Which products are popular in different countries

Technologies

Tools: Visual Studio Code, Jupyter Notebook, Power BI, Tableau, PowerPoint

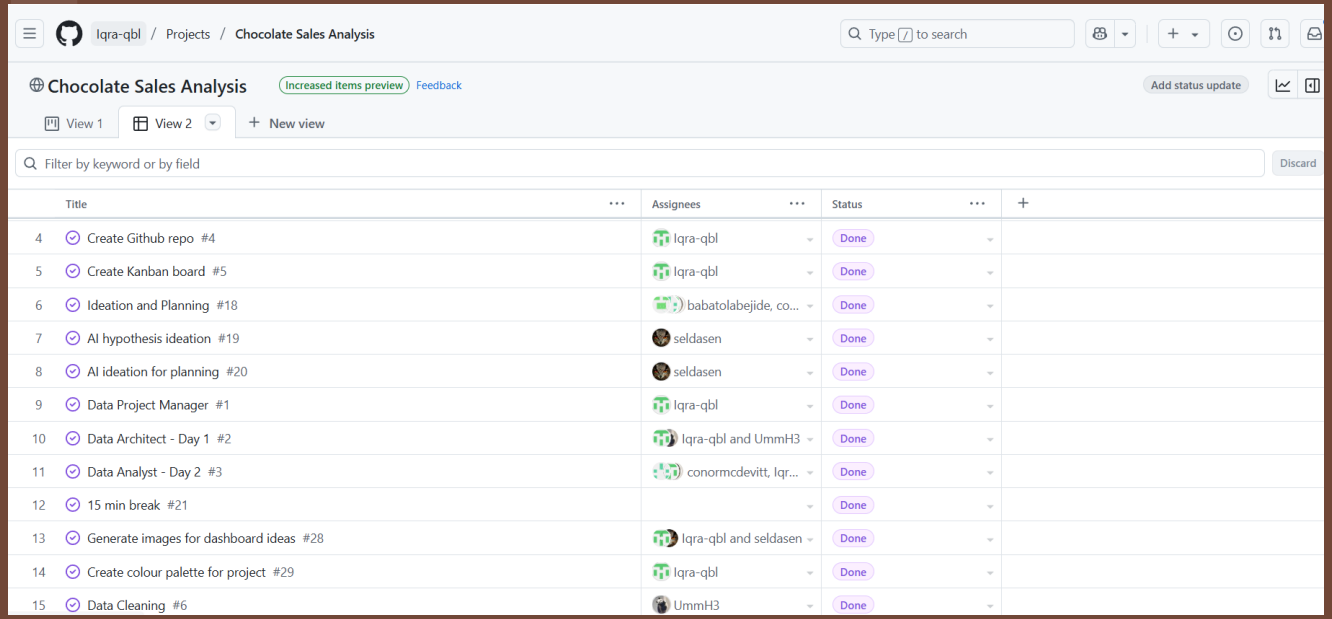
Wireframing: Balsamic Wireframes

Project Management: GitHub Projects, Google Meets

Version Control: GitHub for collaboration

Libraries & Frameworks: Python (Pandas, NumPy, Plotly, Seaborn,Pingoiun), Power BI, Tableau

Project Board



Assigned Tasks: 33

MoSCoW Prioritisation:

Must have: interactive dashboards: Tableau & PowerBi and proper documentation

Should have: simple and followable code

Could have: nicely stylized code and dashboard

Would have: sales forecast

Dashboard Wireframe

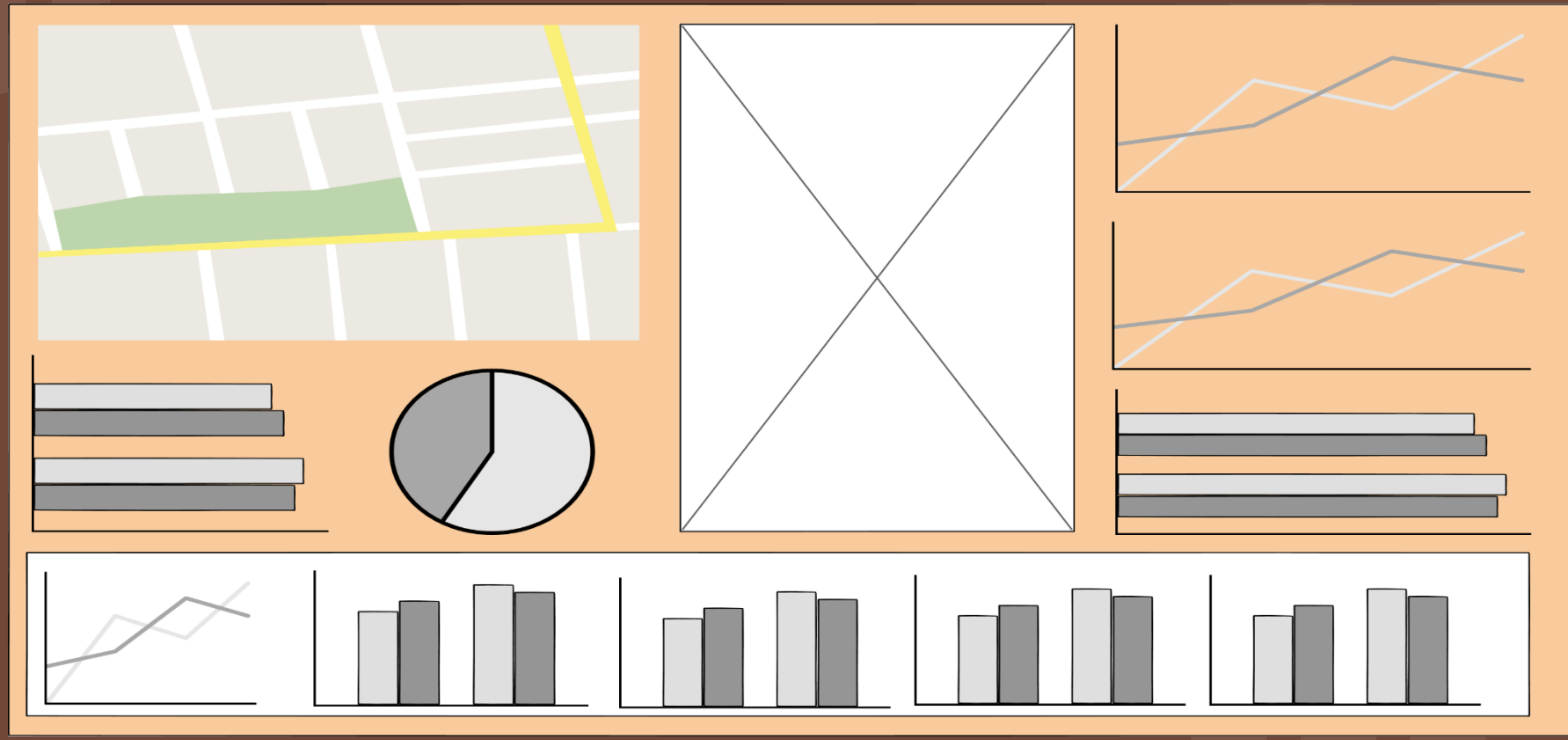


Tableau Features

1. **Interactive**: countries are filterable in map, the months and product category are filterable in chocolate sales line chart and bar chart

2. Hovering over any chart will further explain the data specific e.g. regarding total revenue, product/product category, country etc

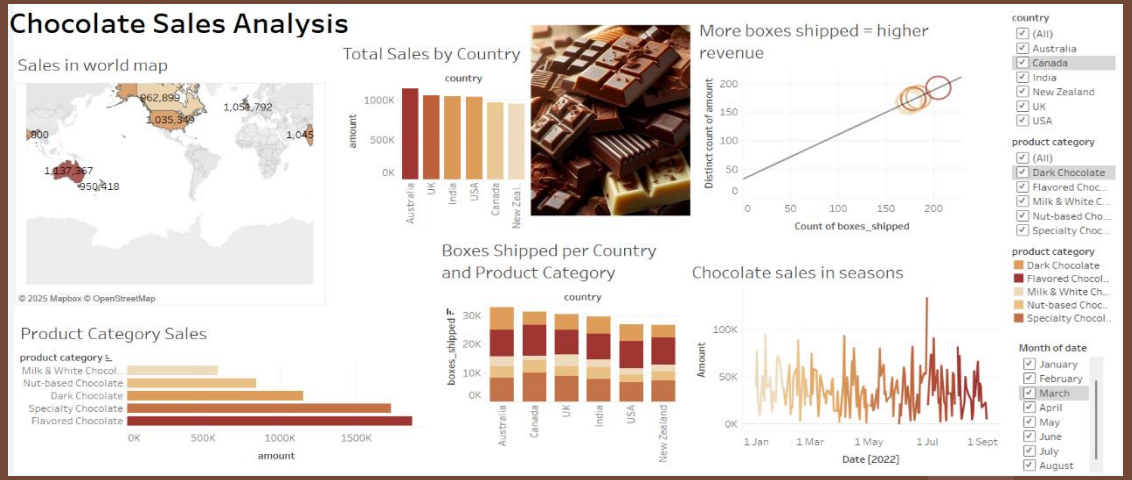
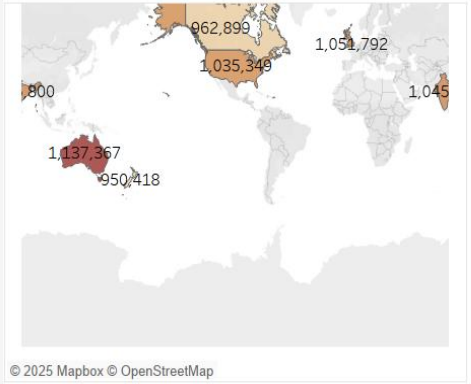


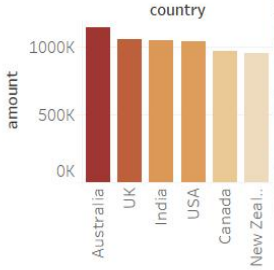
Tableau Features

Chocolate Sales Analysis

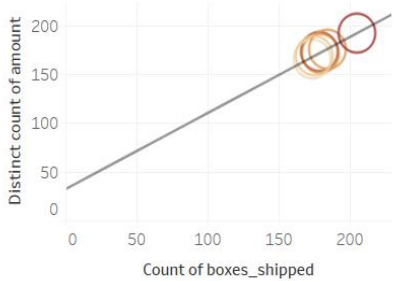
Sales in world map



Total Sales by Country



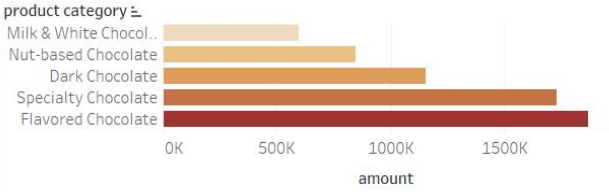
More boxes shipped = higher revenue



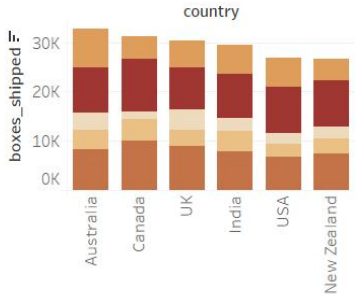
country
- (All)
- Australia
- Canada
- India
- New Zealand
- UK
- USA

product category
- (All)
- Dark Chocolate
- Flavored Choc...
- Milk & White C...
- Nut-based Cho...
- Specialty Choc...

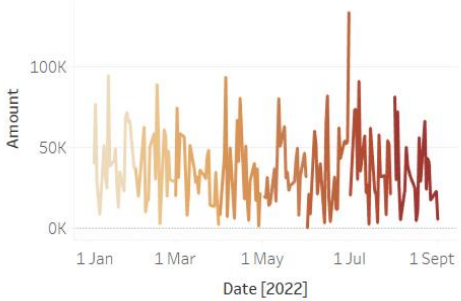
Product Category Sales



Boxes Shipped per Country and Product Category



Chocolate sales in seasons



product category
- Dark Chocolate
- Flavored Chocol..
- Milk & White Ch..
- Nut-based Choc..
- Specialty Chocol..

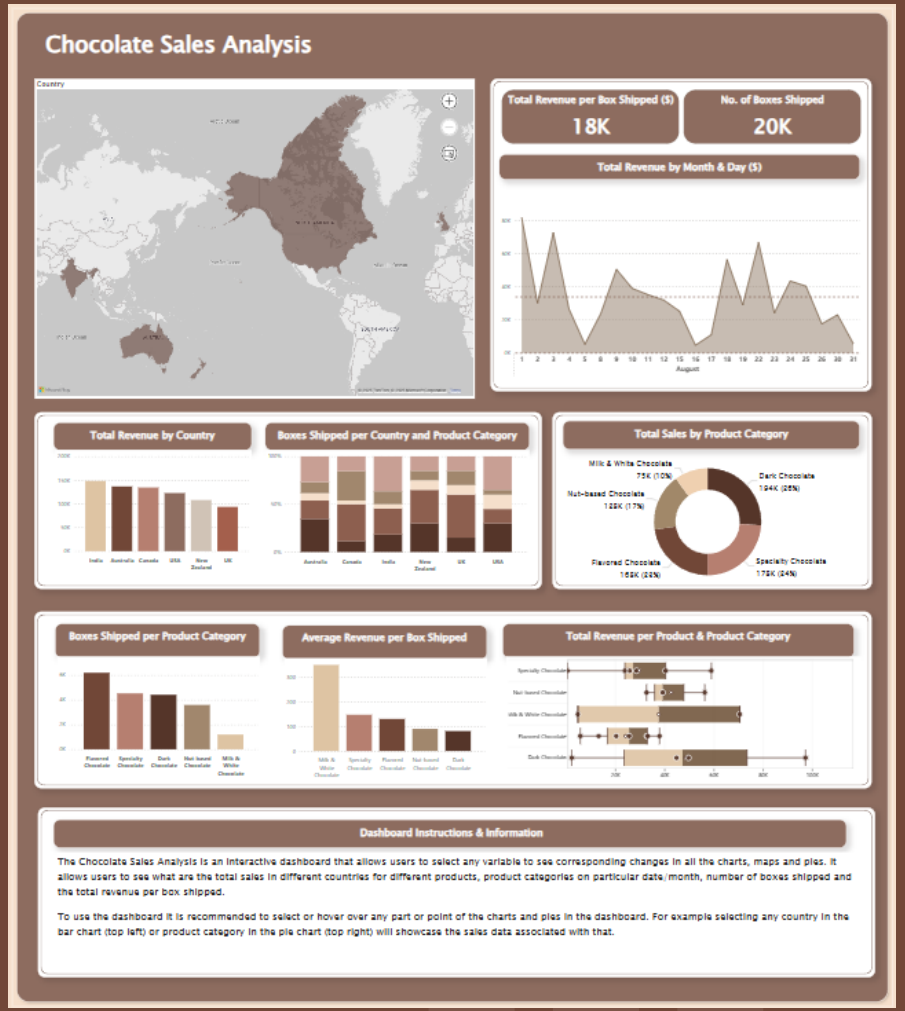
Month of date
- January
- February
- March
- April
- May
- June
- July
- August

PowerBi Features

1. **Interactive**: all variables i.e. total revenue, country, total sales by category, boxes shipped, product category are interlinked so selecting one will show all the relevant data across different charts and map

2. At the bottom there are instructions on using the different features of the dashboard

3. Hovering over any chart will further explain the data specific e.g. regarding total revenue, product/product category, country etc

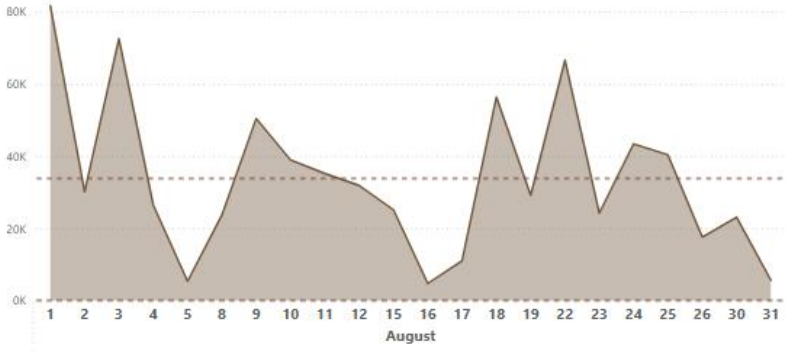


Chocolate Sales Analysis

Total Revenue per Box Shipped ($)
18K

No. of Boxes Shipped
20K

Total Revenue by Month & Day ($)



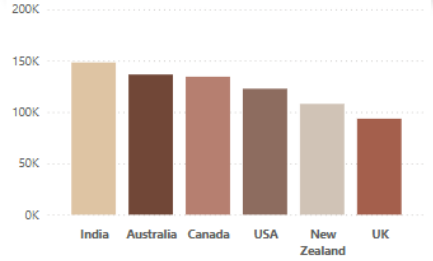
Total Revenue by Country

Boxes Shipped per Country and Product Category

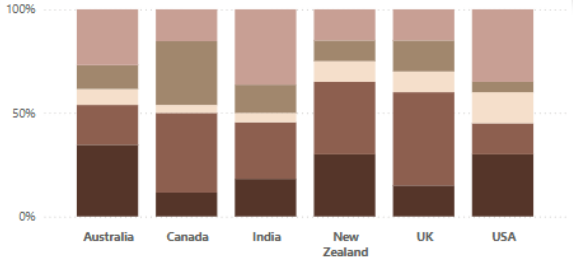
Total Sales by Product Category

August

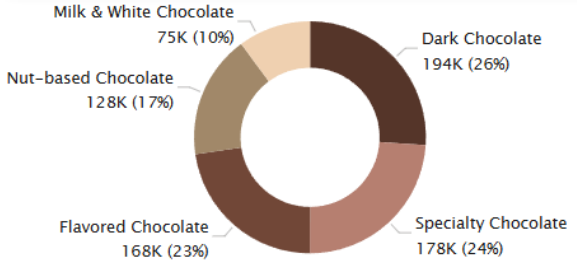
Total Revenue by Country



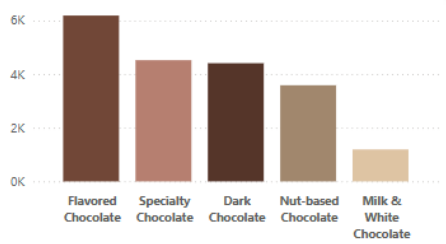
Boxes Shipped per Country and Product Category



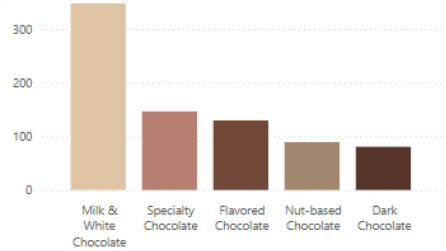
Total Sales by Product Category



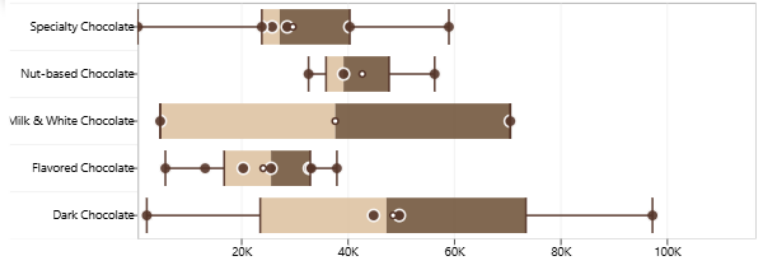
Boxes Shipped per Product Category



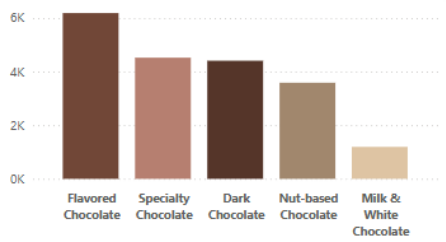
Average Revenue per Box Shipped



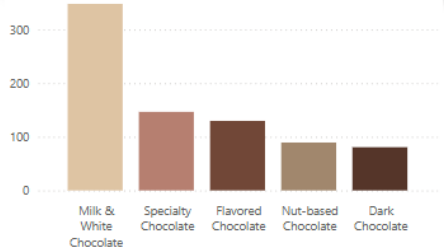
Total Revenue per Product & Product Category



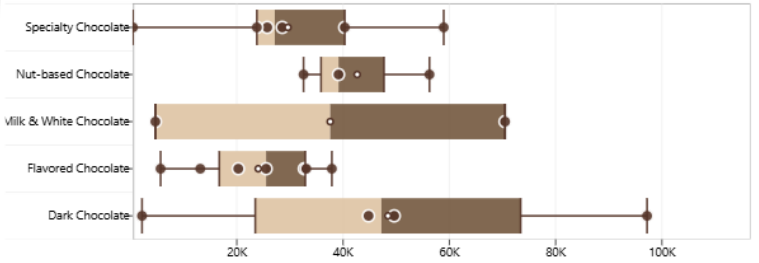
Boxes Shipped per Product Category



Average Revenue per Box Shipped



Total Revenue per Product & Product Category



Dashboard Instructions & Information

The Chocolate Sales Analysis is an interactive dashboard that allows users to select any variable to see corresponding changes in all the charts, maps and pies. It allows users to see what are the total sales in different countries for different products, product categories on particular date/month, number of boxes shipped and the total revenue per box shipped.

To use the dashboard it is recommended to select or hover over any part or point of the charts and pies in the dashboard. For example selecting any country in the bar chart (top left) or product category in the pie chart (top right) will showcase the sales data associated with that.

Chocolate Sales Analysis – Data Analytics Project

Objectives

- Load and preprocess the Chocolate dataset.
- Clean the dataset by removing columns not required,standardise date column,normalise and convert columns where needed.
- Perform exploratory data analysis (EDA) to understand data distribution and relationships.

Inputs

- **Dataset:** kagglehub.dataset_download("atharvasoundankar/chocolate-sales")
- **Required Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Plotly
- **Columns of Interest:** 'Sales Person', 'Country', 'Product', 'Date', 'Amount', 'Boxes Shipped'

Outputs

- **Cleaned dataset:** Processed dataset stored as a CSV file for analysis (`df_cleaned.csv`).

Section 2: Sales Data Analysis

Importing Libraries and Packages

Loading Python packages that we will be using in this project to carry out the analysis. For example Numpy to compute numerical operations and handle arrays, Pandas for data manipulation and analysis, Matplotlib, Seaborn and Plotly to create different data visualisations

[4]
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('whitegrid')
import plotly.express as px
import warnings
warnings.filterwarnings('ignore')
```
Python

Loading the previously cleaned dataset in previous section and then using .head() to check correct dataset has been selected

[5]
```python
df = pd.read_csv("Output\df_cleaned.csv")
df.head()
```
Python

Section 3 : Advance Data Visualisation

Changing work directory

To run the notebook in the editor, the working directory needs to be changed from its current folder to its parent folder. Thus, we first access the current directory with os.getcwd()

```
import os
current_dir = os.getcwd()
current_dir
```

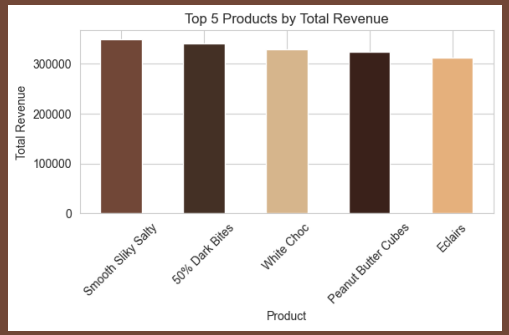
[1] ✓ 0.0s Python

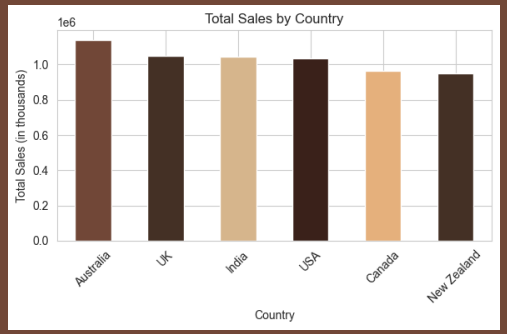
```
's:\\Documents\\Code Institute\\vscode-projects\\Chocolate Sales Analysis\\Chocolate-Sales-Analysis\\jupyter_notebooks'
```

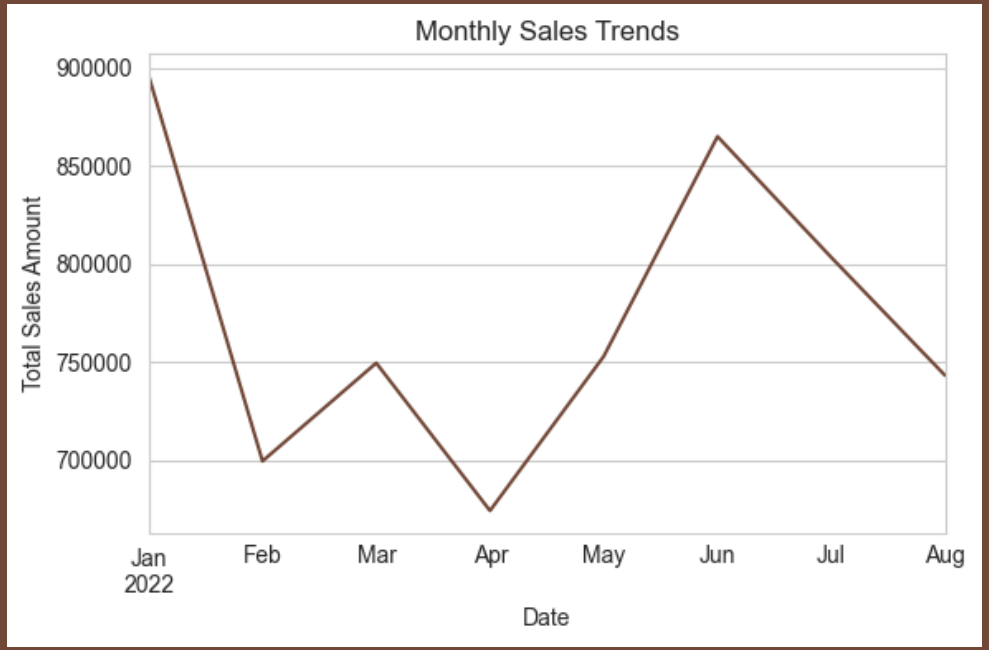
Then we make the parent of the current directory the new current directory by using:

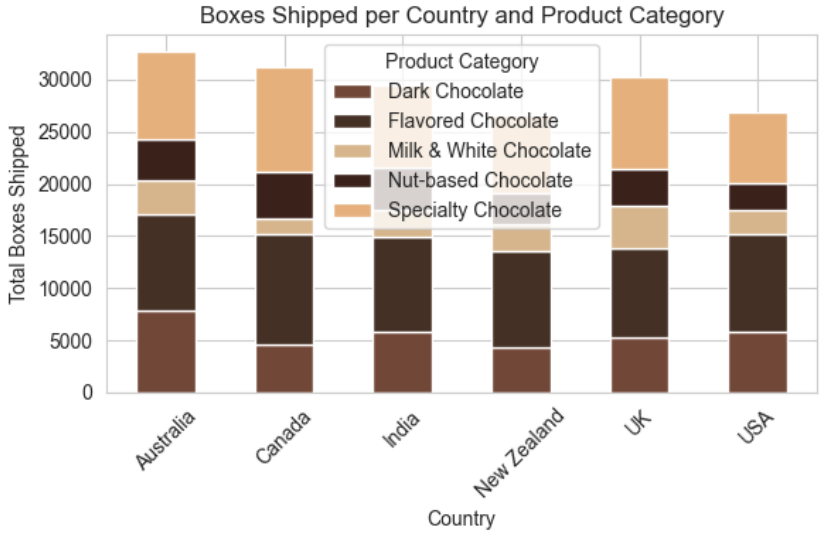
- os.path.dirname() to get the parent directory
- os.chir() to define the new current directory

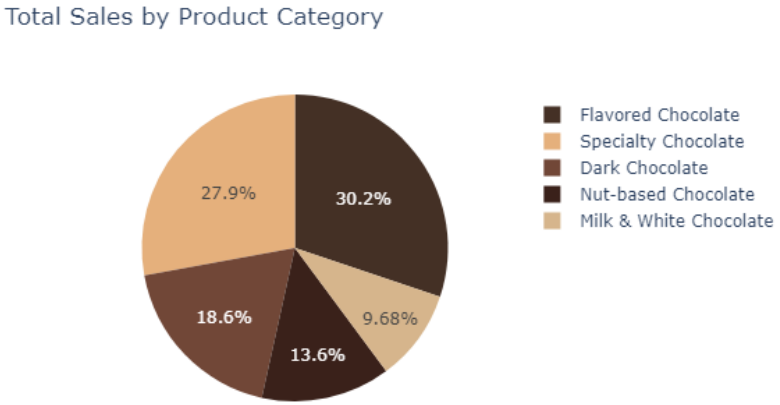
```
os.chdir(os.path.dirname(current_dir))
print("You set a new current directory.")
```


Top 5 Products by Total Revenue
Total Revenue
300000
200000
100000
0
Smooth Silky Salty
50% Dark Bites
White Choc
Peanut Butter Cubes
Eclairs
Product

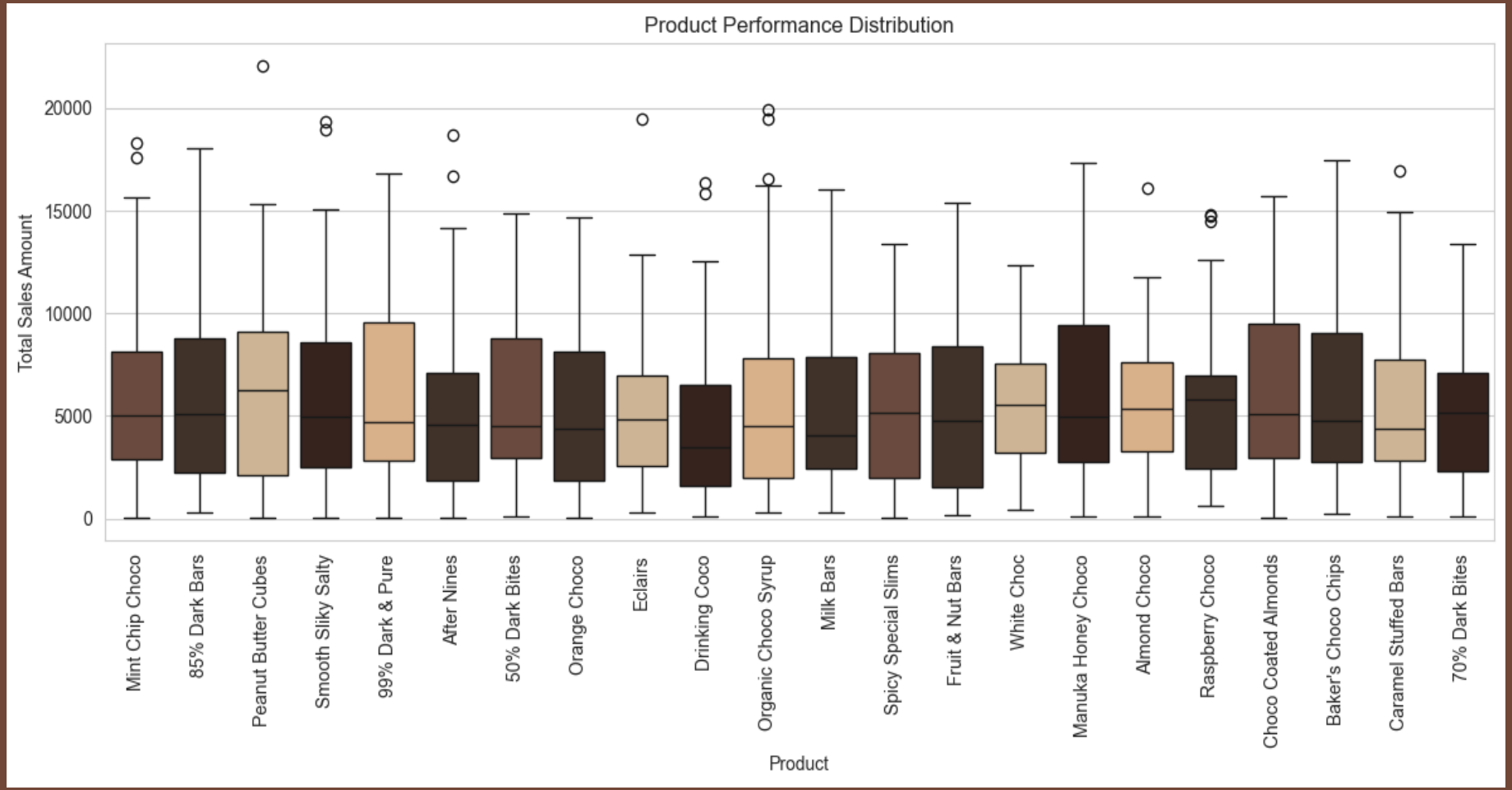

Total Sales by Country
1e6
Total Sales (in thousands)
1.0
0.8
0.6
0.4
0.2
0.0
Australia
UK
India
USA
Canada
New Zealand
Country

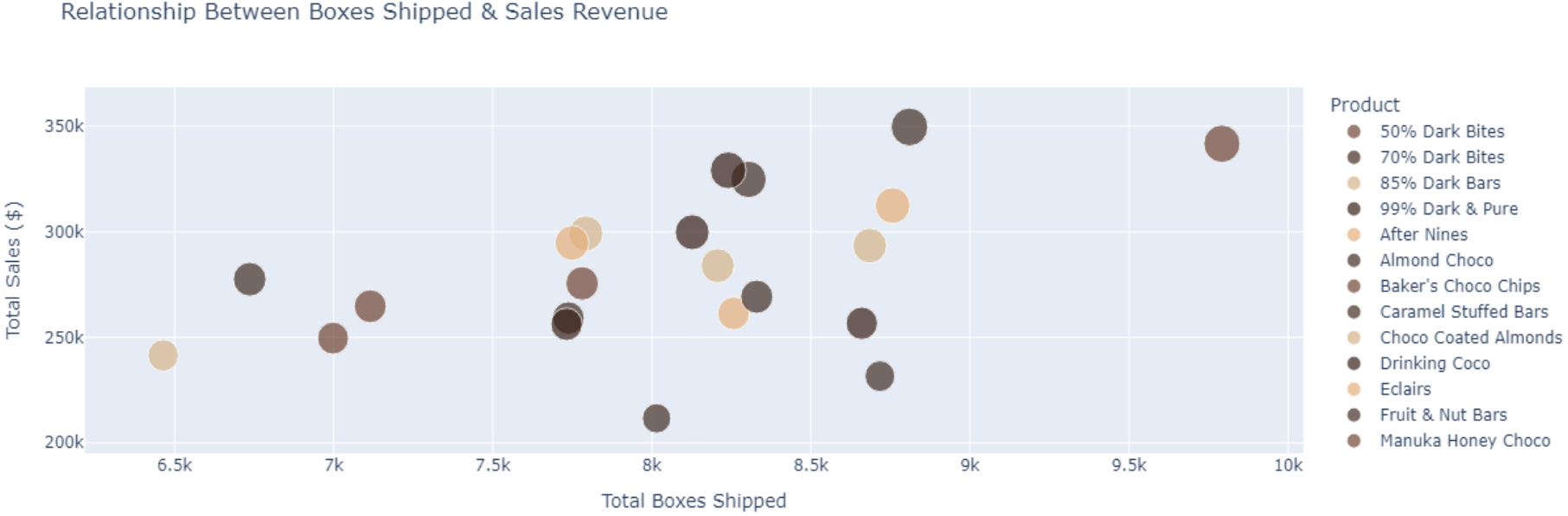

Monthly Sales Trends
Total Sales Amount
900000
850000
800000
750000
700000
Jan
2022
Feb
Mar
Apr
May
Jun
Jul
Aug
Date


Boxes Shipped per Country and Product Category
Product Category
Dark Chocolate
Flavored Chocolate
Milk & White Chocolate
Nut-based Chocolate
Specialty Chocolate
Total Boxes Shipped
30000
25000
20000
15000
10000
5000
0
Australia
Canada
India
New Zealand
UK
USA
Country


Total Sales by Product Category
27.9%
30.2%
18.6%
13.6%
9.68%
Flavored Chocolate
Specialty Chocolate
Dark Chocolate
Nut-based Chocolate
Milk & White Chocolate

Sales Breakdown by Country & Product Category


Product Performance Distribution
Total Sales Amount
20000
15000
10000
5000
0
Mint Chip Choco
85% Dark Bars
Peanut Butter Cubes
Smooth Sliky Salty
99% Dark & Pure
After Nines
50% Dark Bites
Orange Choco
Eclairs
Drinking Coco
Organic Choco Syrup
Milk Bars
Spicy Special Slims
Fruit & Nut Bars
White Choc
Manuka Honey Choco
Almond Choco
Raspberry Choco
Choco Coated Almonds
Baker's Choco Chips
Caramel Stuffed Bars
70% Dark Bites
Product


Relationship Between Boxes Shipped & Sales Revenue
Total Sales ($)
350k
300k
250k
200k
6.5k
7k
7.5k
8k
8.5k
9k
9.5k
10k
Total Boxes Shipped
Product
50% Dark Bites
70% Dark Bites
85% Dark Bars
99% Dark & Pure
After Nines
Almond Choco
Baker's Choco Chips
Caramel Stuffed Bars
Choco Coated Almonds
Drinking Coco
Eclairs
Fruit & Nut Bars
Manuka Honey Choco

Insights & Finding

1. Summary Statistics (Sales & Shipments)

- **Average Sales Amount per Transaction**: $5,652.31
- **Standard Deviation of Sales**: $4,102.44 (indicating high variability)
- **Minimum Sale**: $7.00 | **Maximum Sale**: $22,050.00
- Median Sales Amount: $4,868.50
- Average Boxes Shipped per Transaction: 161.8

2. Correlation Analysis

- Sales Amount vs. Boxes Shipped: -0.0188 (Weak Negative Correlation).

- This suggests that increasing the number of boxes shipped does not strongly impact total revenue. Some high-revenue sales may involve fewer boxes but premium-priced chocolates

3. Top 5 Countries by Total Sales Revenue

- Australia ($1,137,367), UK ($1,051,792), India ($1,045,800), USA ($1,035,349) and Canada – ($962,899)

- Australia leads in total sales, followed closely by UK and India

- India's high sales suggest strong demand despite potentially different market dynamics

4. Top 5 Best-Selling Chocolate Products

- Smooth Silky Salty ($349,692), 50% Dark Bites ($341,712), White Choc ($329,147), Peanut Butter Cubes ($324,842), Eclairs ($312,445)

- Dark chocolate & peanut butter varieties perform well, suggesting a preference for premium or unique flavours

- Smooth Silky Salty is the best-selling product, potentially due to unique taste or marketing strategies

Insights & Finding

5. Monthly Sales Trend (First 7 Months of 2022)

- January 2022: $896,105, February 2022: $699,377, March 2022: $749,483, April 2022: $674,051, May 2022: $752,892
- January has the highest sales (possibly due to post-holiday chocolate purchases)
- Sales drop in February and April, suggesting potential seasonality effects
- Sales increase in June again, maybe due to summer holidays

Collaboration & Outcomes

Outcomes

Are you happy with the final product?
Yes

What do you hope to achieve in the next development cycle?

Create sales forecast for next months

What would you do differently if you could start again?

Maybe select dataset with more months or years recorded data

Development Problems

Problems that arose during development?:

Git collaboration

In group conflicts and resolutions?
No

Did you find any of the behaviour related content useful? Teamwork, problem solving etc?

Yes, mindful collaboration and problem- solving, work division

Interactivity: Overall good

Summary

Overall group dynamic:
Good, friendly, professional

Overall satisfaction: 9/10 (-1 sales forecast)

What we learned:
Working with new team members

Our experiences: We had good troubleshooting and mindset for the project

Q&A