

A circular botanical illustration border surrounds the central text. It features various plants including ferns, a red leaf, a green leaf, a yellow flower, a purple flower, and a branch with small pink flowers.

DATA ANALYST PROJECT CASE STUDY

by: Padmasini Venugopal



INTRODUCTION

DAIRY PRODUCTS CONSUMPTION IN INDIA (2019-2022)

India is the world's largest milk producer with almost 22% of global production. To understand the sales pattern to improvise the pricing strategies among different brands this analysis has been performed.





Project Overview



DATA AND PROJECT

It's an open-source data set from Kaggle.

- Data source: [link](#)
- Dataset: [source](#)

DURATION

It took almost 2 to 4 weeks for me to complete the entire analysis.

TOOLS USED

- Python
- Jupyter notebook
- Excel
- Tableau

TECHNIQUES APPLIED

- Exploratory visual analysis
- Created correlation matrices and heatmaps
- Created a choropleth map in Python
- Regression analysis (supervised machine learning)
- Cluster analysis (unsupervised machine learning)
- Time series analysis

```
jupyter Dairy data cleaned Last Checkpoint: 07/02/2023 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel) O
+ -> Run Code

In [1]: # Importing Libraries
import pandas as pd
import numpy as np
import os

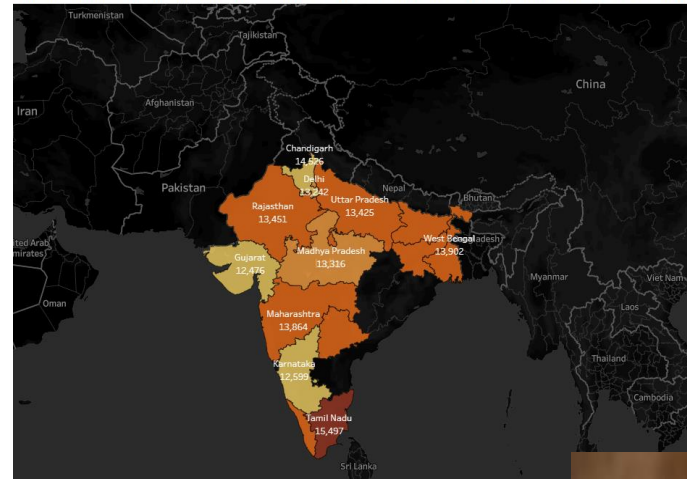
In [2]: path = r"C:\Users\harip\ Dairy data Analysis\"

In [3]: # Importing the dataset
df = pd.read_csv(os.path.join(path, '02 Data', 'Original data', 'dairy_dataset.csv'), index_col = False)

In [4]: # Determining the data
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4325 entries, 0 to 4324
Data columns (total 23 columns):
 #   Column              Non-Null Count  Dtype
---  --
 0   Location            4325 non-null   object
 1   Total Land Area (acres)  4325 non-null   float64
 2   Number of Cows      4325 non-null   int64
 3   Farm Size           4325 non-null   object
 4   Date                4325 non-null   object
 5   Product ID          4325 non-null   int64
 6   Product Name        4325 non-null   object
 7   Brand               4325 non-null   object
 8   Quantity (liters/kg)  4325 non-null   float64
 9   Price per Unit      4325 non-null   float64
10  Total Value         4325 non-null   float64
11  Shelf Life (days)  4325 non-null   int64
12  Storage Condition   4325 non-null   object
13  Production Date     4325 non-null   object
14  Expiration Date     4325 non-null   object
15  Quantity Sold (liters/kg)  4325 non-null   int64
16  Price per unit (sold)  4325 non-null   float64
17  Approx. Total Revenue (INR)  4325 non-null   float64
18  Customer Location   4325 non-null   object
19  Sales Channed       4325 non-null   object
20  Quantity In Stock (liters/kg)  4325 non-null   int64
21  Minimum Stock Threshold (liters/kg)  4325 non-null   float64
22  Reorder Quantity (liters/kg)  4325 non-null   int64
dtypes: float64(8), int64(5), object(10)
memory usage: 277.3+ KB
```

THE MAP SHOWS THE DAIRY PRODUCTION IN DIFFERENT STATES OF INDIA AND THEIR AVERAGE REVENUE PER DAY



THE PROCESS

1. THE DATA PREPARATION PHASE

To streamline sales data for dairy products in India, I conducted a comprehensive scan to identify and eliminate duplicates, missing figures, and other anomalies. Subsequently, I removed extraneous columns and duplicates from the dataset.

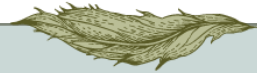
2. THE ANALYSIS PHASE

Proceeding with my further investigations I performed an exploratory visual analysis, created correlation matrices and choropleth maps to understand the geography, and additionally conducted regression and cluster analysis.

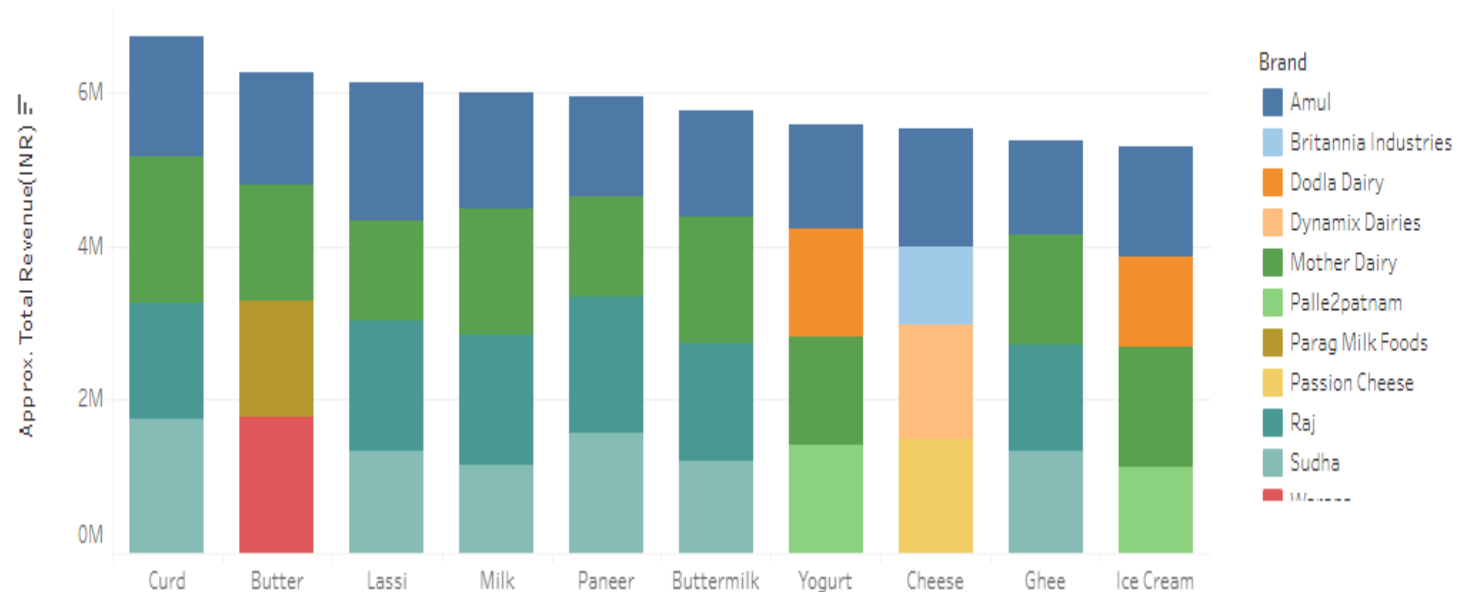
3. THE RESULTS

Finally, after a thorough analysis the final presentation was created using Tableau. The stakeholders were advised based on the results of the analysis with a clear disclaimer regarding the limitations of the data.

EXPLORATORY ANALYSIS

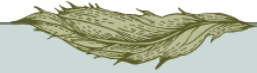


TOTAL REVENUE BASED ON BRANDS AND PRODUCT NAMES



- The most sold product which contributes to the highest sales revenue is **Curd** followed by it is **butter and lassi**.
- Amul, Mother Dairy, and Raj are the top three brands with the highest revenue.

INVENTORY MANAGEMENT

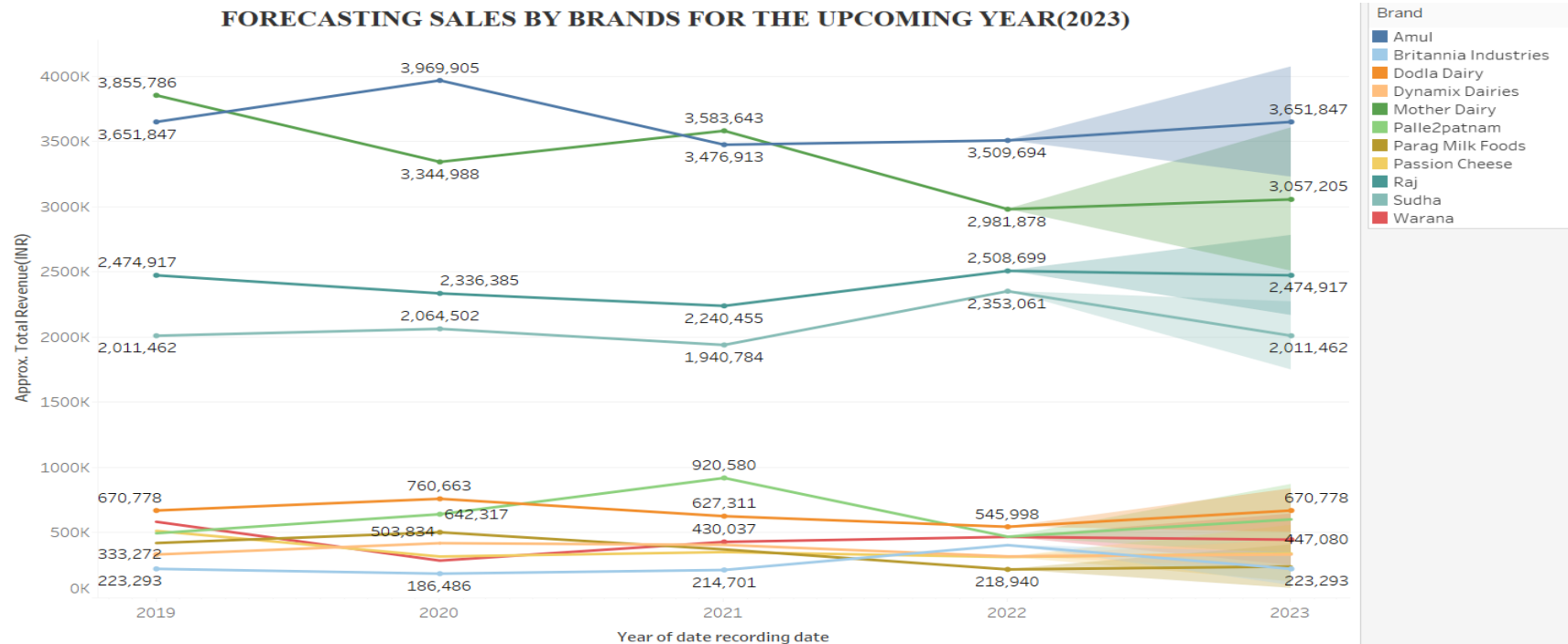
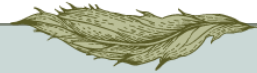


DETERMINING THE SHELF LIFE OF STOCKS



The clustering plot between "shelf life" and "quantities in stock" shows that **milk products in stock** have a shelf life of 0 to 40 days.

FORECAST FOR 2023



Based on the forecast predictions, there may be a **mild increase in sales** in most of the dairy brands including the top brands like Amul and Mother Dairy by 2023 in India.

RETROSPECTIVE



WHAT WENT WELL?

I was surprised to discover a new tool I was unaware of all these years. Python just made the work so much easier when compared to all the other tools especially when the data is larger in numbers.

CHALLENGES FACED?

At first, learning Python was a huge challenge for me. It was difficult to learn a new programming language and remember all the necessary code. However, with consistent practice, I have gained some hands-on experience. Although I am not yet perfect, I am still practicing.

RECOMMENDATIONS

- It is recommended to expand sales into other states and union territories.
- We should create more marketing campaigns for moderately performing brands in order to improve their sales revenue.
- As we live in an online generation, it would be beneficial to promote more online methods for purchase and delivery.
- India has many regional languages, so it would be wise to advertise in the regional language of each area, which can help customers better understand the products.



Thank you



The Link to the Final Presentation

[Click here](#)

For further queries kindly contact
drpadmasini.v@gmail.com