SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, TRICHY

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

DATA SCIENCE [21CSS303T]

PROJECT REPORT

April 2025

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, TRICHY

DATA SCIENCE PROJECT REPORT

NAME	E.G.PRADEEP V.KOUCIKAN
REG NO	RA2211004050026 RA2211004050041
SEMESTER	VI
PROGRAME	B.TECH ECE
ACADEMIC YEAR	2024-2025

ABSTRACT

This project explores the application of data science techniques to analyze the launch prices of Samsung mobile phones over the past five years (2020–2024). By utilizing Python's capabilities for data analysis, this study focuses on identifying pricing trends, flagship and mid-range segmentation, and predictive insights for future pricing. Through systematic statistical operations and exploratory data analysis, the project highlights how simple computational methods can provide valuable strategic intelligence in the mobile phone industry.

INTRODUCTION

The smartphone industry generates vast amounts of data every year, reflecting changes in consumer preferences, technological advancements, and pricing strategies. Analyzing this data can help manufacturers like Samsung tailor future devices to market demands. This project focuses on the Samsung mobile phones launched from 2020 to 2024 and applies basic data science techniques using Python to extract meaningful pricing trends and insights.

OBJECTIVE

- Data Collection and Preparation: Create a dataset containing Samsung mobile model names, launch years, and prices.

- Exploratory Data Analysis (EDA): Investigate model launches per year, price averages, flagship/mid-range categorization.

- Predictive Analysis: Forecast the average price for 2025 based on historical trends.

- Insights Generation: Provide a strategic view of Samsung's pricing evolution.

IMPORTANCE OF DATA VISUALIZATION

Data visualization plays a crucial role in this project by enabling easier interpretation of pricing distributions and trends. Visualizations such as bar charts and line graphs can help identify patterns in pricing strategy, mid-range vs flagship ratios, and predict future pricing behavior. Visual insights support quicker, data-driven decision-making.

DATASET DESCRIPTION

The dataset includes details of Samsung smartphones launched from 2020 to 2024. Each record contains the model name, launch year, and initial launch price (in USD).
Key Attributes:
1. Model Name
2. Launch Year
3. Launch Price (USD)
Samsung Mobile Phones Dataset:
Galaxy S20 (2020) - $999
Galaxy Note 20 (2020) - $1199
Galaxy S21 (2021) - $799
Galaxy Z Fold 3 (2021) - $1799
Galaxy S22 (2022) - $899
Galaxy Z Flip 4 (2022) - $999
Galaxy S23 (2023) - $999
Galaxy Z Fold 5 (2023) - $1899
Galaxy S24 (2024) - $1099
Galaxy Z Flip 6 (2024) - $1099

DATA PROCESSING

The data was processed using Python. Lists were created for models, launch years, and prices. Standard analysis was performed using dictionary manipulations, Counter, and basic Python functions. No missing values were found as the dataset was prepared manually.

OPERATIONS PERFORMED USING THE DATASET

1. **Total number of models launched each year**

Python Code:

```
from collections import Counter
launch_data = [2020, 2020, 2021, 2021, 2022, 2022, 2023, 2023, 2024, 2024]
models_per_year = Counter(launch_data)
print(models_per_year)
```

2. **Average launch price per year**

Python Code:

```
from collections import defaultdict
```

```
prices = [999, 1199, 799, 1799, 899, 999, 999, 1899, 1099, 1099]
years = [2020, 2020, 2021, 2021, 2022, 2022, 2023, 2023, 2024, 2024]
year_price = defaultdict(list)
for year, price in zip(years, prices):
   year_price[year].append(price)
average_price_per_year = {year: sum(p)/len(p) for year, p in year_price.items()}
print(average_price_per_year)
```

3. Overall average price across all five years

Python Code:

```
overall_avg_price = sum(prices) / len(prices)
print(overall_avg_price)
```

4. Highest priced model per year

Python Code:

```
model_names = ["Galaxy S20", "Galaxy Note 20", "Galaxy S21", "Galaxy Z Fold 3",
           "Galaxy S22", "Galaxy Z Flip 4", "Galaxy S23", "Galaxy Z Fold 5",
           "Galaxy S24", "Galaxy Z Flip 6"]
year_model_price = list(zip(years, model_names, prices))
highest_price_per_year = {}
for year in set(years):
   year_models = [(model, price) for y, model, price in year_model_price if y == year]
   highest = max(year_models, key=lambda x: x[1])
   highest_price_per_year[year] = highest
print(highest_price_per_year)
```

5. Year with the most expensive Samsung phone launched

Python Code:

```
most_expensive = max(zip(model_names, prices, years), key=lambda x: x[1])
print(most_expensive)
```

6. Categorize models into Mid-Range (<1000 USD) and Flagship (>=1000 USD)

Python Code:

```
mid_range = [model for model, price in zip(model_names, prices) if price < 1000]
flagship = [model for model, price in zip(model_names, prices) if price >= 1000]
print("Mid-Range:", mid_range)
```

```
print("Flagship:", flagship)
```

7. Percentage of Mid-Range vs Flagship models

Python Code:

```
mid_range_count = len(mid_range)
flagship_count = len(flagship)
total = mid_range_count + flagship_count
mid_range_percent = (mid_range_count / total) * 100
flagship_percent = (flagship_count / total) * 100
print("Mid-Range %:", mid_range_percent)
print("Flagship %:", flagship_percent)
```

8. Predicted price for 2025 (assuming 5% price increase)

Python Code:

```
last_avg_2024 = sum([1099, 1099]) / 2
predicted_2025 = last_avg_2024 * 1.05
print(predicted_2025)
```

9. Insights

Python Code:

```
print("""
Samsung maintains a strong flagship segment while offering mid-range devices.
Foldable phones from 2021 onward significantly raised average launch prices.
Galaxy Z Fold 5 in 2023 was the highest priced model.
""")
```

CONCLUSION

Through this project, we demonstrated how simple data science techniques can uncover important insights into Samsung's smartphone pricing strategy. The analysis revealed trends in average pricing, segmentation into mid-range and flagship devices, and provided a forecast for future model pricing. These insights can guide strategic planning for both companies and consumers in understanding market movements and pricing expectations.