CS7DS4 / CSU44065 Data Visualization 2025-26 Assignment 3

Student Name: Mihir Apte Student No: 25334927

DECLARATION: "I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at http://tcd ie.libguides.com/plagiarism/ready-steady-write."

Global Smartphone Landscape 2025: An Interactive Visualization

1. Dataset Description and Motivation

The dataset used in this project contains information about smartphones released in 2025 across multiple brands. Each record represents a mobile phone model and includes attributes such as price (USD), RAM, storage, camera resolution, battery capacity, display size, charging speed, operating system, processor, 5G support, user rating, and release month and year. The dataset is multivariate and heterogeneous, combining numerical, categorical, binary, and temporal attributes.

This dataset is complex because meaningful insights cannot be obtained from a single static chart. Price, for example, is influenced by multiple factors such as brand, hardware specifications, and release timing. Exploring relationships such as price versus performance, trade-offs between battery and camera quality, or differences between brands and operating systems requires interactive and multi-view visualization. Visualization is therefore essential to reduce cognitive load and allow users to explore patterns, comparisons, and trends effectively.

2. Tools and Technologies

The visualization was implemented using Python with the Dash framework and Plotly for interactive graphics. Pandas was used for data preprocessing and transformation. These tools were chosen because they support interactive, web-based visualizations and allow fine-grained control over visual encoding and layout.

3. Data Complexity and Analytical Tasks

The complexity of the dataset arises from both its dimensionality and the relationships between attributes. There are many quantitative variables that interact with each other, such as price, RAM, battery capacity, and camera megapixels, along with categorical variables like brand and operating system. Temporal attributes further add complexity by allowing analysis over time.

The visualization is designed to support several key analytical tasks. Users can compare smartphones across brands and operating systems, explore relationships between price and technical specifications, identify outliers such as unusually expensive or feature-rich devices, and analyse how prices vary across different release periods. These tasks would be difficult to perform using tables or static plots, especially given the number of attributes involved.

4. Visualization Design and Idioms Used

To address the dataset's complexity, a multi-idiomatic interactive dashboard was designed using Plotly Dash in Python. The main visualization combines several complementary views within a single screen layout.

The primary view is an interactive scatter plot that shows price on the x-axis and a user-selected attribute (such as camera megapixels, RAM, or battery capacity) on the y-axis. Additional visual channels are used to encode more information: colour represents brand groups, and point size represents another selectable attribute such as display size. This allows multiple variables to be explored simultaneously without overwhelming the user.

To complement this, a reduced scatterplot matrix (SPLOM) is included to show relationships between a carefully selected subset of key variables: price, RAM, camera resolution, battery capacity, and rating. The matrix was deliberately simplified to avoid clutter and visual overload, focusing only on attributes that provide meaningful comparative insight.

A faceted view is also provided using small multiples, where price and rating relationships are shown separately for each major brand. This design choice allows direct comparison across brands while maintaining clarity and consistency across views.

5. Interaction, Faceting, and Temporal Representation

Interaction plays a central role in this visualization. Users can filter the dataset by brand, operating system, price range, and 5G support using interactive controls. These filters update all visualizations simultaneously, enabling linked exploration across multiple views.

Faceting is used in the brand comparison view, where each brand is displayed in its own panel. This makes it easier to compare trends without overplotting and supports perceptual comparison across categories.

For temporal analysis, a static time-based chart is used to show how average smartphone prices change across release months. A static temporal view was chosen instead of animation to ensure that trends remain clearly visible and interpretable at all times, especially given the single-year scope of the dataset. This approach avoids unnecessary complexity while still satisfying the requirement for temporal analysis.

6. Data Pre-processing and Design Decisions

Before visualization, several preprocessing steps were applied. Release month names were normalised and converted into numeric values to support time-based aggregation. A combined month-year field was created where possible to support temporal grouping. Brands were grouped into the top six categories, with all others combined into an “Other” category to reduce visual clutter and improve readability in faceted views.

Design decisions were guided by the goal of balancing richness with clarity. The scatterplot matrix was intentionally limited to fewer dimensions to prevent overcrowding. Consistent colour encoding was used across all views to maintain visual coherence. Interactive filtering was prioritised over adding additional static charts, as it allows users to control the level of detail based on their interests.

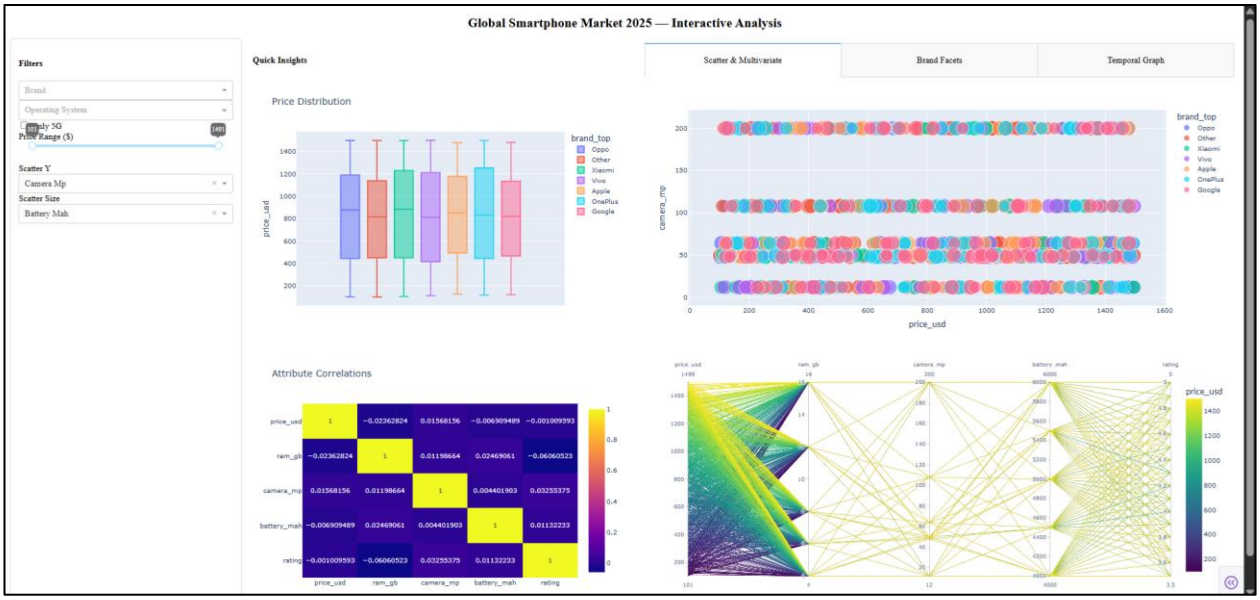
7. Evaluation, Strengths, and Limitations

The final visualization effectively addresses the complexity of the dataset by combining multiple idioms, interaction, and faceting within a single coherent dashboard. Users can explore trade-offs between price and features, compare brands, and observe temporal patterns without being overwhelmed. The design is flexible and supports multiple analytical tasks, aligning well with the goals of exploratory data analysis.

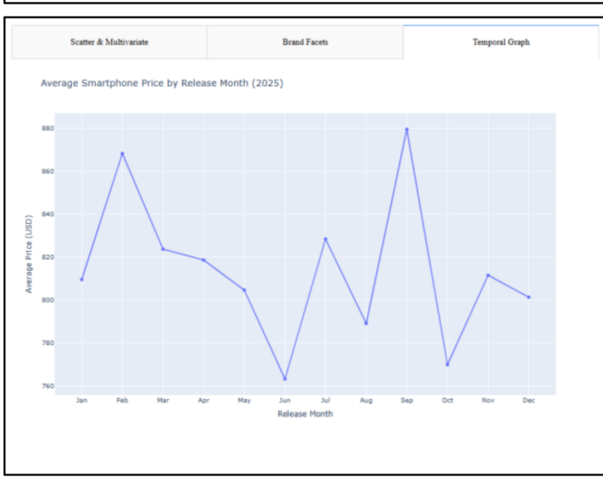
However, there are some limitations. The dataset covers only a single year, which restricts deeper long-term temporal analysis. Additionally, some advanced visualizations, such as hierarchical views, were explored but not included due to data sparsity and stability issues. Despite these limitations, the visualization successfully demonstrates how interactive, multi-view designs can reduce complexity and support meaningful insight.

8. Conclusion

This project demonstrates how interactive, multi-idiom visualization can be used to address the complexity of a real-world, multivariate dataset. By combining thoughtful visual design with interaction and preprocessing, the visualization enables meaningful exploration of global smartphone data and supports a range of analytical tasks


Global Smartphone Market 2025 — Interactive Analysis
Filters
Brand
Operating System
Only 5G
Price Range ($)
Scatter Y
Camera Mp
Scatter Size
Battery Mah
Quick Insights
Scatter & Multivariate
Brand Facets
Temporal Graph
Price Distribution
brand_top
Oppo
Other
Xiaomi
Vivo
Apple
OnePlus
Google
price_usd
camera_mp
Attribute Correlations
price_usd
ram_gb
camera_mp
battery_mah
rating
1
−0.02362824
0.01568156
−0.006909489
−0.001009593
0.01198664
0.02469061
−0.06060523
0.004401903
0.03255375
0.01132233


Scatter & Multivariate
Brand Facets
Temporal Graph
Oppo
Other
Xiaomi
Vivo
Apple
OnePlus
Google
brand_top
rating
price_usd


Scatter & Multivariate
Brand Facets
Temporal Graph
Average Smartphone Price by Release Month (2025)
Average Price (USD)
Release Month
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec