

Capstone Project:

StreamLit Project - Laptop Prices Prediction

Name: Rutuja Patil

NUID: 002827468

GitHub: <https://github.com/PatilRutu66/StreamLit-Project>

Introduction:

The application describes how a machine learning-based model for predicting laptop prices is being developed with the goal of changing how merchants and customers interact with the laptop industry. The main goal is to accurately forecast laptop pricing by utilizing a variety of features and specs, enabling customers to make knowledgeable purchases. This model offers merchants insightful information on pricing tactics that can be used to improve customer satisfaction and profitability by streamlining inventory management and customizing marketing campaigns.

Abstract:

An abstract outlining the dataset used for the project opens the notebook. Twelve different columns in this dataset, which includes 1273 computers, list variables including the name of the manufacturer, the type of laptop, the screen size and resolution, the CPU and RAM specs, the storage capacity, the GPU characteristics, the operating system, and the price in Indian rupees. It draws attention to how useful the dataset is for assessing laptop market trends, which helps retailers and manufacturers plan their price and product offerings. Customers can also benefit from this study by using it to make well-informed judgments about what to buy based on hardware characteristics and price.

Dataset Description:

The structure and contents of the dataset are briefly summarized, highlighting the breadth and depth of information that may be used to analyze laptop configurations and costs. The dataset consists of 1273 data points spread across 12 different attributes.

Data Preparation and Loading:

Libraries and Data Loading: The notebook lists several Python libraries used for data handling and analysis, including NumPy, Pandas, and visualization tools like Matplotlib and Seaborn. Additionally shown is the basic code for loading the dataset from a CSV file. We started by utilizing Pandas, a potent Python data processing package, to load the dataset. The dataset, which was obtained from Kaggle, has 1273 entries with 12 attributes that are pertinent to laptop pricing and specifications.

Data Cleaning

Data cleaning focused on ensuring accuracy and completeness:

Removing Duplicates: We checked for and removed any duplicate records to prevent skewing our analysis and model predictions.

Handling Missing Values: We identified columns with missing data and employed appropriate imputation strategies to fill these gaps, depending on the nature and distribution of the data.

Correcting Data Types: Certain columns were converted to more appropriate data types to facilitate analysis. For example, price values were converted from strings to numerical data types to enable mathematical operations.

Data Exploration

Exploratory data analysis (EDA) aimed to uncover patterns, detect anomalies, and test hypotheses using statistical figures and visualization tools:

Analytical Statistics

To provide an overview of the distribution's shape, central tendency, and dispersion, we offered descriptive statistics. This covered the price, which was the main variable of interest, in particular, and included means, medians, ranges, and standard deviations.

Visual representation

Box Plots and Histograms: To make sure our model trains on accurate and representative data, we used box plots to identify outliers in the data and histograms to comprehend the price distribution.

Correlation Matrix: To show which characteristics have the most influence on laptop pricing, a correlation matrix was created to show the links between various features and the price.

Scatter Plots: These provide a visual representation of the possible linear or non-linear correlations between important characteristics, such as RAM and price.

Author: @Rutuja Patil

GitHub: <https://github.com/PatilRutu66>

StreamLit application for Laptop Price Predictions

- This prediction application estimates laptop prices based on their features and specifications.



Laptop Brand & Type

Company

Apple

Type

Ultrabook

Weight (KG)

0.92

Screen Specs

Touch Screen

Yes

Panel Type

IPSPanelRetinaDisplay

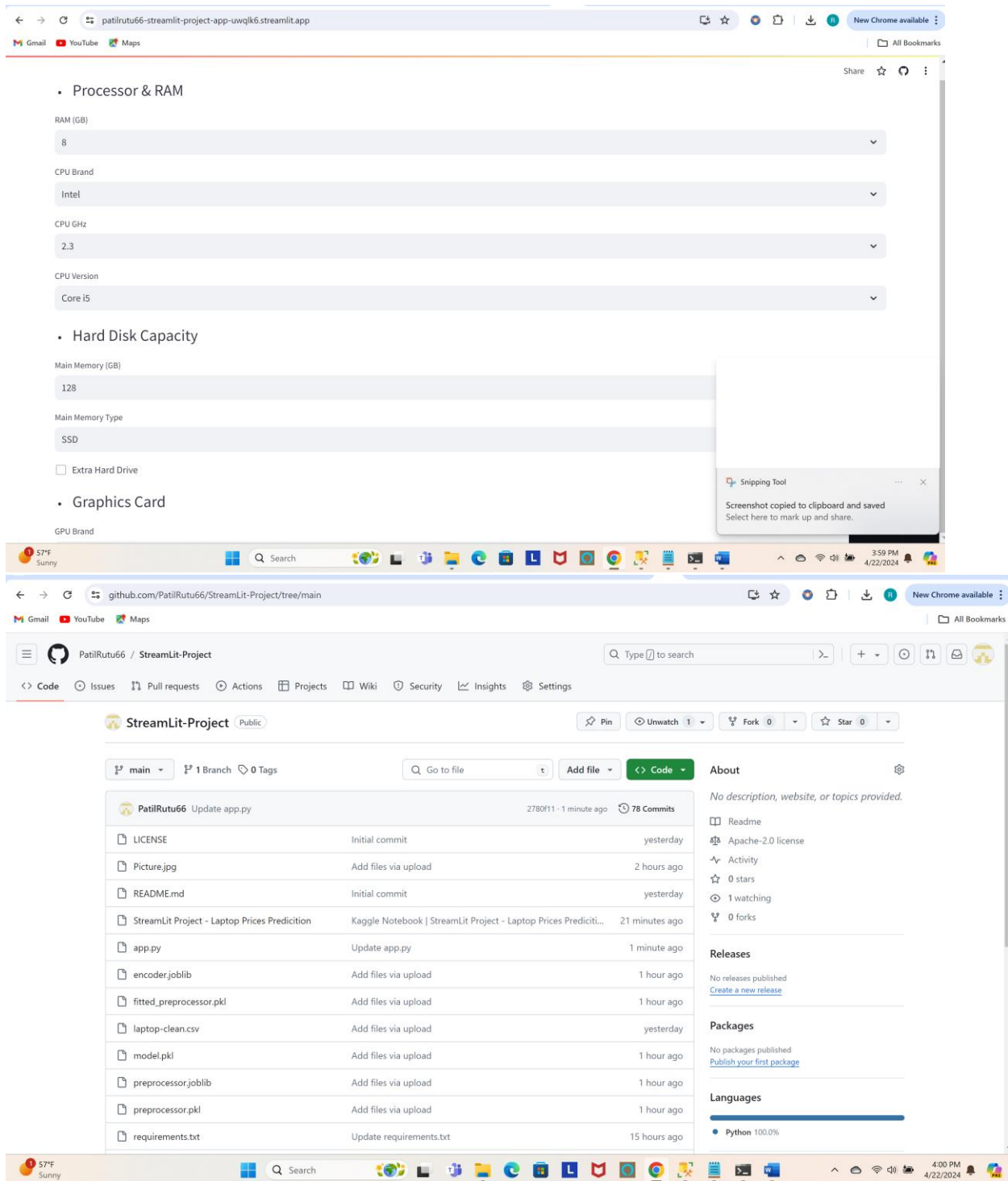
Resolution

2560x1600

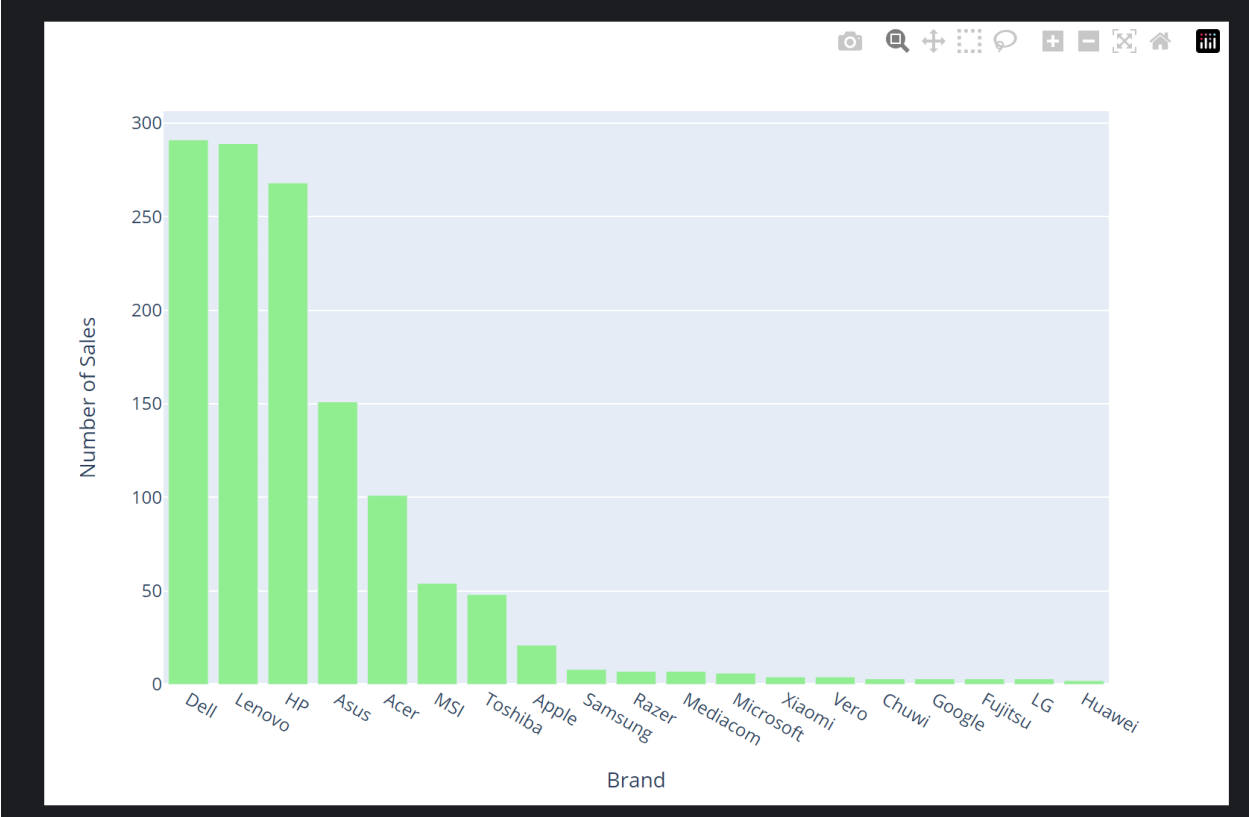
Inches

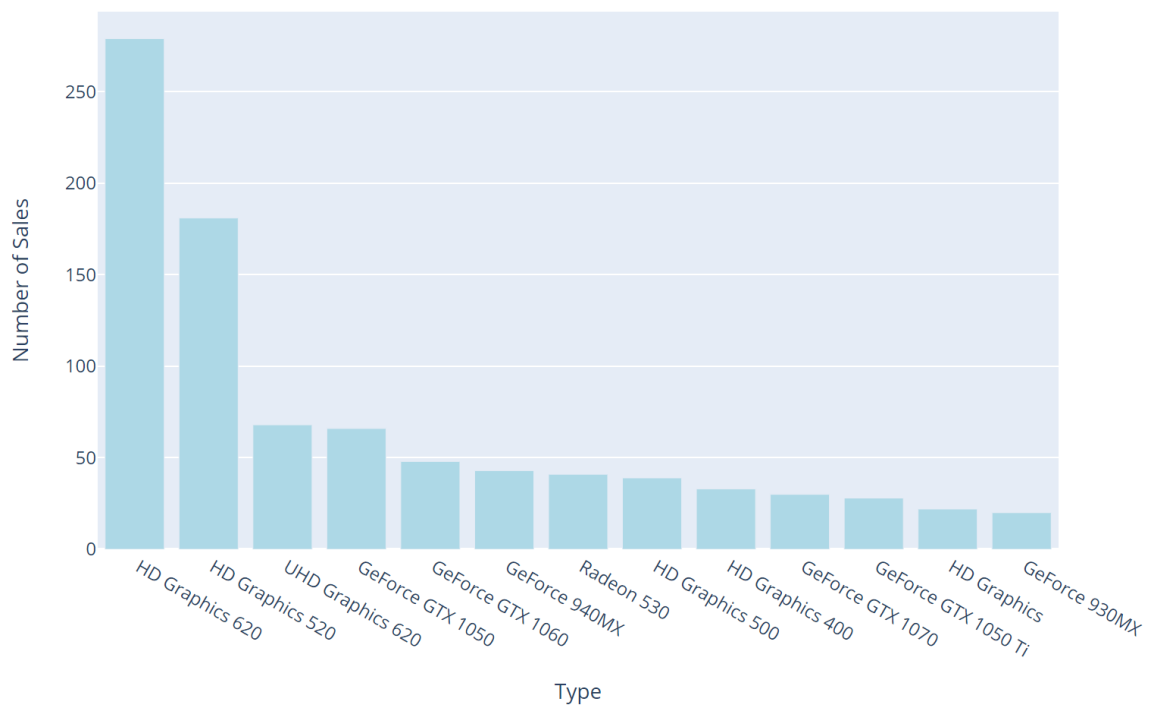
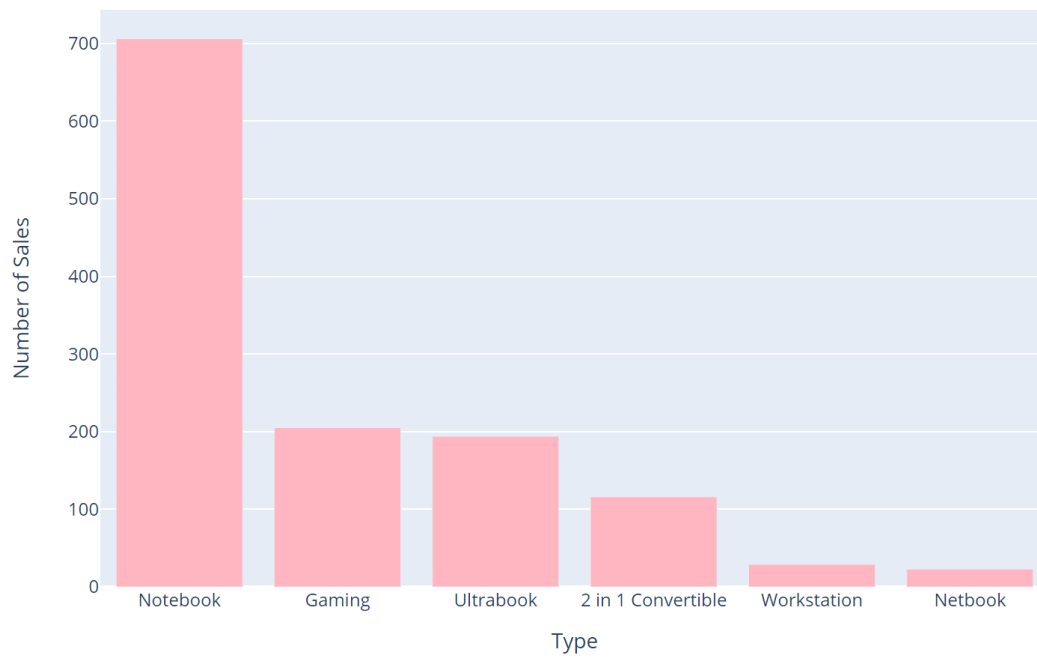
11.60

Processor & RAM

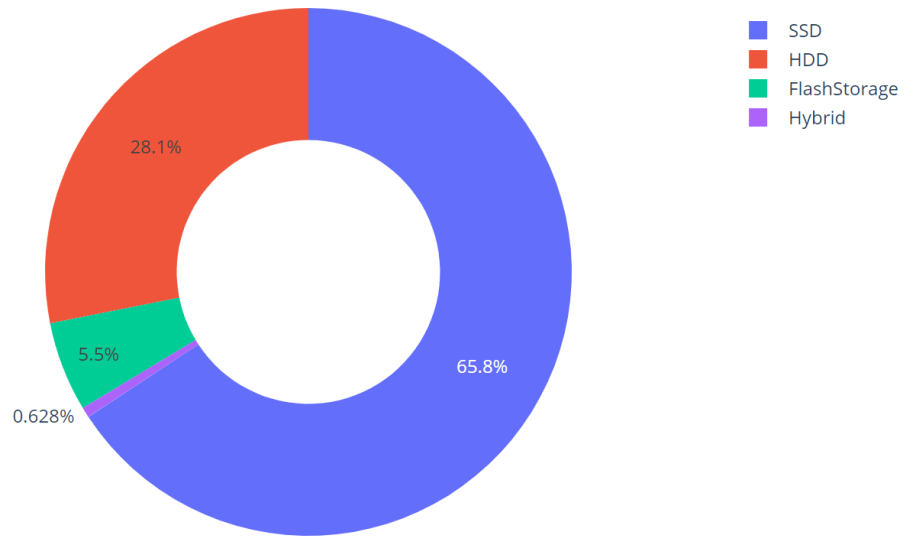


Few snapshots of the data analysis:

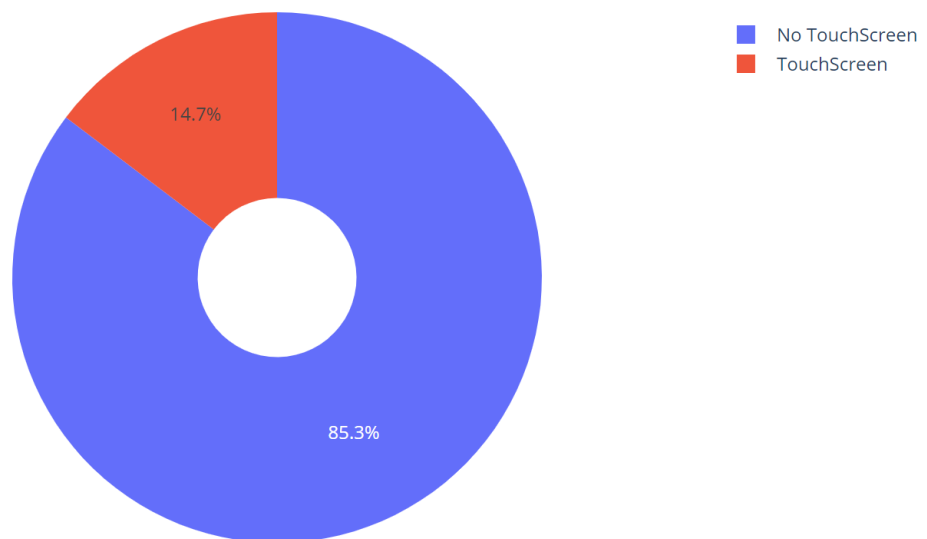




Main Laptop Memory Type



Touch Screen Or Not





My StreamLit app link: <https://patilrutu66-streamlit-project-app-uwqlk6.streamlit.app/>

Conclusion:

Using a dataset of laptop pricing and characteristics, I have created a thorough laptop price prediction model for this project. The model was trained using advance machine learning techniques and then integrated into an easy-to-use Streamlit application to provide users with real-time laptop pricing estimates based on their inputs.

Future Work:

In order to increase the project's applicability and performance, I want to expand it to incorporate predictive analysis for additional electronic devices. It may be possible to investigate potential increases in model accuracy and performance by conducting additional research and improving ensemble learning and deep learning methodologies.

References:

<https://www.kaggle.com/competitions/kschool-master-data-science-streaming-2023>
<https://www.kaggle.com/code/ahmetdzdar/streamlit-machine-learning-app>
<https://docs.streamlit.io/get-started/installation>

<https://docs.streamlit.io/deploy/streamlit-community-cloud/manage-your-app>
<https://docs.streamlit.io/>

License:

Apache 2.0 open source license