PROPOSAL

Linear Regression

Of

By
Reem Alsuqaih
Njood AL Qahtani

Abstract:

Who doesn't know Amazon? Amazon is one of the most popular and largest shopping sites, Free shipping on millions of items. Enjoy low prices and great deals on the largest selection of everyday essentials and other products, including fashion, home, beauty, electronics, Alexa Devices, sporting goods, toys, automotive, pets, baby, books, video games, musical instruments, office supplies, and more. We chose the best-selling products in the electronics section and applied linear regression to them.

Design

This project is one of the T5 Data Science Boot Camp requirements. Data provided by Amazon.sa. In this module, we will be laying the foundation for our analysis by processing and exploring a large amount of data on electronic sales. The dataset contains information regarding the best electronic sales of various type of electronic. *Get the data* here

Learning Objectives

After this module, we should be able to do the following:

1- Use panda's library to analyze different features of the dataset, which includes,

2- Use Beautiful Soup and request library to get the data.

3- Use web scrapping for web pages that will analyze it

4- Read the dataset.

5- Use plotly library to visualize the given results

6- Plot graphs like bar graphs and pie charts

Understanding the dataset

This dataset contains a list of electronic items around 1000 product. In this data set there are 7 columns. Their names and data types as follows:

Fields include

- Name of product
- Rate of product
- Price of product
- Review

Algorithms

- Web scrapping for amazon web page
- Request the data by request library
- Cleaning the data and remove null values
- Transform the string data to numbers.

Tools

- Pandas for data manipulation
- Beautiful Soup library
- Request library
- Matplotlib for plotting
- Plotty for visualizations
- World cloud

Communication

- The slides will be provided here, Feel free to any pull requests.