

## *Web Wonder Women (WWW)*

*Rozina Ahmed Reda*  
*202400970*

*Menna Mohamed Hassan*  
*202400939*

### Phase 1:

- Used Ebay API to collect data about electronics. The data collected included title, **price**, image url, item id, item url. The type of products information collected was diverse. For example, headphones, keyboard, smartphone, air conditioner and some famous Tech companies products such as Lenovo and Dell.
- Then, saved all data in one dataset CSV file (electronics.csv) using Pandas library.
- Cleaning: The price column included all values in a form of integer and currency (ex. 49.99 USD) so, we split it into two columns, price column holds the integer and currency column holds “USD” .
- Pre-processing: The value of price column is saved as datatype of “object” (string) so, we converted its data type to integer.
- Analysis: Used Pandas to create a bar graph that shows the price frequency. From this graph, we conclude that 500\$ is the most common price in our electronics dataset.

- Analysis: Created a new dataset to include the product titles that were found at least three times in the dataset. This dataset includes the product title and its frequency. Then, created a horizontal bar graph the number of occurrences of most common products in our dataset. We conclude that Apple Macbook is the most popular product in the dataset.
- Used Selenium to collect data about products in categories laptops, phones, watches. The collected data include title, link, image url, and **rating(count)**.
- Then, saved the data in one dataset as an excel file (xlsx)
- Graph Construction: Used this dataset and NetworkX library to visualize the network edges of products with the number of ratings for each product. This was done by calculating the betweenness centrality to know the central node of our dataset and have it visualized with different color.

## Phase 2:

- Used Selenium to collect data for motors products category. The data includes title, link, image url, **rating(value)**, and **price**. And added **rating(value)** column for electronics dataset.
- Then, saved all of motors data in one dataset
- Cleaning: the price column in motors include all values in a form of integer and currency (ex. \$200.00) so, we replaced it with a new column that includes the integer value of price only.

- Pre-processing: The value of price column is saved as datatype of “object” (string) so, we converted its data type to float.
- Then, saved it to motors\_cleaned.csv file
- NetworkX:
  - Created a graph to visualize the relation between products in **motors** category and their prices. The product with the highest is shown in different color. Concluded that Odyssey Battery is the product with the highest price in our motors dataset.
  - Created a graph to visualize the relation between products in **electronics** category and their prices. The product with the highest is shown in different color. Concluded that Apple Macbook is the product with the highest price in our electronics dataset.
- Heatmap: Created a heatmap to visualize the relation of prices in electronics and motors datasets
- Analysis:
  - Created a line graph to show the **prices** of electronics and motors.
  - Created a line graph to show the **ratings** of electronics and motors.

## Phase 3:

- 3D Analysis: Created a graph to show the relation between electronics products attributes of Price, Rating, and number of reviews (ratings).

- UI: Created a search engine that starts by asking the user whether he wants to search in motors or electronics category. Second, it asks the user to enter a product name or brand that the user wants to search for. Third, it asks the user whether he wants to know the price, rating, or the review number. Finally, it shows the rows that have the product or brand name and the required attribute the user searched for.