<https://github.com/Rami200574/RamiSaaifan20220099Ass5>

**Project: eBay Tech Deals Scraper and Analysis**

**Methodology**

This project involves two main tasks:

1. **Web Scraping**:
   * A **Python scraper** was developed using **Selenium** to extract data from the eBay Tech Deals page (<https://www.ebay.com/globaldeals/tech>).
   * The scraper collects key product details, including:
     + **timestamp**: The date and time the data was scraped.
     + **title**: The title of the product.
     + **price**: The discounted price of the product.
     + **original\_price**: The original price (if available).
     + **shipping**: Shipping information for the product.
     + **item\_url**: The product's URL on eBay.
   * The scraper runs every three hours using **GitHub Actions**, building a robust dataset over a period of two days.
2. **Data Cleaning**:
   * The raw data scraped by the scraper was cleaned in the **clean\_data.py** script:
     + The **price** and **original\_price** columns were cleaned to remove the "US $" currency symbol and commas.
     + Missing **original\_price** values were replaced with the corresponding **price**.
     + The **shipping** column was cleaned to ensure consistent information.
     + The **discount\_percentage** column was calculated and rounded to two decimal places.
3. **Exploratory Data Analysis (EDA)**:
   * In the **EDA.ipynb** Jupyter notebook, several analyses were conducted to gain insights into the dataset:
     + **Time Series Analysis**: Analyzed the number of deals per hour.
     + **Price and Discount Analysis**: Explored the distribution of prices and discounts.
     + **Shipping Information**: Examined the frequency of shipping options.
     + **Text Analysis on Product Titles**: Analyzed the occurrence of specific keywords in the titles.
     + **Price Difference Analysis**: Calculated and visualized the absolute price difference between original and discounted prices.
     + **Top Deals**: Displayed the top 5 products with the highest discounts.

**Key Findings from EDA**

1. **Time of Day Patterns**:
   * The number of deals showed variation across different hours of the day. Certain hours of the day had a higher volume of deals, indicating possible marketing strategies or timing preferences.
2. **Price Distribution**:
   * The distribution of product prices was skewed, with many products falling into a lower price range. This indicates that eBay offers a significant number of affordable tech deals.
3. **Discount Analysis**:
   * A significant portion of products had substantial discounts, with some items offering discounts exceeding 50%. This insight could help identify the most discounted categories or brands.
   * The analysis also revealed that most products had discounts ranging from 10% to 40%.
4. **Shipping Information**:
   * Most products provided standard shipping options, with a small portion offering free or expedited shipping. This could be important for understanding customer preferences and competitive shipping strategies.
5. **Text Analysis on Product Titles**:
   * Popular keywords such as "Apple", "iPhone", and "Laptop" appeared frequently in product titles. This suggests that these are some of the top product categories for tech deals on eBay.
6. **Price Difference**:
   * The absolute price difference between the original price and the discounted price varied widely, with some items offering significant savings.
7. **Top Deals**:
   * The top 5 deals with the highest discounts were primarily in the tech gadget category, highlighting the importance of high discounts to attract customers.

**Challenges Faced**

1. **Dynamic Content Loading**:
   * eBay's page uses lazy loading for displaying products, which required implementing scrolling functionality to ensure that all products were loaded before scraping.
2. **Handling Missing Data**:
   * Some products did not have an "original\_price" available. To handle this, I replaced missing values with the discounted "price", ensuring a complete dataset for further analysis.
3. **Scraper Stability**:
   * The scraper faced occasional issues with loading elements due to the dynamic nature of eBay's page. Using WebDriverWait and ensuring proper scrolling resolved these issues.
4. **Data Inconsistencies**:
   * The data collected from the eBay page was not always consistent (e.g., missing shipping info, varying formats for price). These were handled during the cleaning phase.

**Potential Improvements**

1. **Handling More Product Variants**:
   * The scraper currently captures only basic product details. Future improvements could include scraping additional data like product ratings, reviews, and seller information to provide a more comprehensive dataset.
2. **Optimizing Scraper Speed**:
   * The scraper could be optimized by reducing the number of page loads (e.g., by using headless browsers) or focusing only on certain categories rather than scraping all deals.
3. **Advanced Data Analysis**:
   * Further analysis could be performed to identify trends in product types, sales periods, and discounts. Machine learning models could be implemented to predict future discount trends or recommend the best time to buy specific products.
4. **Handling Multiple Pages**:
   * Currently, the scraper only works with the current page. Implementing functionality to handle multiple pages or navigating through pagination would allow for a more complete dataset.
5. **API Integration**:
   * Instead of scraping HTML data, integrating with eBay’s API (if available) would improve the data collection process and reduce dependency on parsing web page elements.

**Conclusion**

This project successfully automates the process of scraping and analyzing eBay’s tech deals, providing valuable insights into product pricing, discounts, shipping information, and popular product categories. The workflow is set to run automatically every three hours, building a dynamic dataset that can be used for further analysis, trends, and predictions.

The findings from the exploratory data analysis offer actionable insights for businesses or consumers interested in understanding eBay’s pricing and sales patterns, especially in the tech category.