#### Project Description

This Project was been organized in scope of Data Science course at Mtqi Zark,

This document offers a comprehensive overview of the web scraping results obtained from hh.ru, a popular website containing job listings related to C++ and programming positions in various companies. The primary aim of this documentation is to provide detailed insights into the data structure, variables, and potential use cases for the scraped data.

**Author**:

students: Dr. Hans Hofmann

mentors:   
**Data** **Source**: *Website*: hh.ru (Job listings related to programming positions)- scrapped by selenium

**Dataset Description**

This dataset classifies people described by a set of attributes as good or bad credit risks.

This dataset comes with a cost matrix:

Size of dataset:   
Volume: 10 Mb

Colulmns: 10

Rows: 20000

It is worse to class a customer as good when they are bad (5), than it is to class a customer as bad when they are good (1).

### Attribute Description

1. **Title**: The title of the job listing, which succinctly represents the job role.
2. **Company**: The name of the company offering the job, allowing for company-specific analysis.
3. **City**: The location where the job is based, essential for geographical insights.
4. **Experience**: The required experience level for the job, often expressed as a range (e.g., "1-3 years"), aiding in the analysis of skill demand.
5. **Salary**: The salary information, either as a range or a specific salary offered for the position. This is crucial for salary-related investigations.
6. **Rate**: The rating of the job (if available), providing an initial impression of job quality.
7. **Percent**: The rating percentage of the job (if available), offering a more granular view of job satisfaction.
8. **Rate\_count**: The count of ratings received (if available), indicating the popularity and level of interest in the job.
9. **Work\_time**: The work time or schedule for the job (e.g., "Full-time, remote"). This helps understand the work arrangements for different positions.

**Features Description**

| **Feature Name** | **Type** | **Distinct/Missing Values** |
| --- | --- | --- |
|  |  |  |

**Laptop CSV File Documentation**

**Variable Types and Descriptions**

1. **Company**: Text
   * Description: The name of the company that manufactures the laptop.
2. **Model**: Text
   * Description: The model name or number of the laptop.
3. **Display Height**: Numeric (Integer)
   * Description: The height resolution of the laptop's display in pixels.
4. **Display Width**: Numeric (Integer)
   * Description: The width resolution of the laptop's display in pixels.
5. **Display Size**: Numeric (Float)
   * Description: The diagonal size of the laptop's display in inches.
6. **Display Quality**: Text
   * Description: Information about the quality of the display (e.g., Full HD, 4K, etc.).
7. **Display Type**: Text
   * Description: The technology used for the display (e.g., IPS, OLED, etc.).
8. **Processor Name**: Text
   * Description: The name of the processor (CPU) used in the laptop.
9. **Processor Generation**: Text
   * Description: The generation of the processor (e.g., Intel Core i5 10th Gen, Intel Core i7 11th Gen, etc.).
10. **Memory**: Numeric (Integer)
    * Description: The internal storage capacity of the laptop in GB.
11. **Ram**: Numeric (Integer)
    * Description: The RAM (Random Access Memory) capacity of the laptop in GB.
12. **Videocard**: Text
    * Description: The type or name of the integrated graphics card.
13. **Videocard Storage**: Text
    * Description: Information about the storage related to the graphics card.
14. **Illuminated keyboard**: Text
    * Description: Indicates whether the laptop's keyboard is illuminated (yes/no or true/false).
15. **Price**: Numeric (Integer)
    * Description: The price of the laptop in a certain currency (e.g., dollars).
16. **Website index**: Numeric (Integer)
    * Description: An index or identifier for the laptop on a website or platform.
    * 1-notebookCentre.am
    * 2-nout.am
    * 3-buylaptop.am
    * 4-magicnotebooks.am(generated store using scraped data before)

**Laptop Analysis ideas**

1. **Summary Statistics**:
   * Calculate summary statistics for the numerical columns like Display Height, Display Width, Display Size, Memory, RAM, Price, etc. This could include measures like mean, median, standard deviation, and range.
2. **Top Manufacturers and Models**:
   * Identify the top manufacturers and models based on the frequency of appearance in the dataset.
3. **Display Size vs. Price Analysis**:
   * Create a scatter plot to visualize how display size relates to the laptop's price. This can help you understand if larger displays tend to be more expensive.
4. **Processor Analysis**:
   * Group the data by processor names or generations and calculate average prices, memory, RAM, etc. This can help you see which processors are commonly used and their corresponding characteristics.
5. **Graphics Card Analysis**:
   * Explore the distribution of different graphics cards and their associated memory capacities. Visualize this data to understand the prevalence of various graphics card options.
6. **Keyboard Illumination and Price Relationship**:
   * Investigate whether laptops with illuminated keyboards tend to have higher or lower prices. You could use a box plot or bar chart to display this relationship.
7. **Correlation Analysis**:
   * Calculate correlations between numerical variables (e.g., Display Size, Memory, RAM, Price) to identify any potential relationships or patterns.
8. **Company Market Share**:
   * Calculate the market share of each company based on the number of laptop models they have in the dataset.
9. **Processor Generation Trends**:
   * Plot the distribution of processor generations to observe trends in popularity and usage over time.
10. **Comparison of Different Attributes**:
    * Compare attributes like Memory, RAM, and Price across different manufacturers or models to identify trends or differences.
11. **Website Index Analysis**:
    * Investigate whether there's any relationship between the website index and other variables like price or specifications.