# Data Analysis Project

## Data Analyst: addoune salah eddine

## Client/Sponsor: ADDOUNE Co.

## Purpose:

*Write a brief description of why this project is happening below. Why is this project happening? What are the goals?*

* The store manager wants to predict the laptop's price in Algeria so first ask me to analyze the laptop's data .

## Scope / Major Project Activities:

*What are the major parts of this project? List out the high-level steps, activities, or stages of the project, and give a brief description for each.*

| Activity | Description |
| --- | --- |
| data collecting | collect price data from various algerian resources,other info can get it from any resource |
| data cleaning | Cleanse and preprocess the collected data to ensure consistency and accuracy. This includes handling missing values, removing duplicates, standardizing formats, and resolving inconsistencies |
| perform EDA | Perform EDA on the laptop data to gain insights into the market in Algeria. Explore statistical summaries, visualize data distributions, and identify patterns and trends. Analyze the relationships between different variables and their impact on laptop prices. |
| Feature Selection and Engineering | Identify the key features and attributes that significantly influence laptop prices in Algeria. Conduct feature selection techniques to determine the most relevant variables for the predictive model. Additionally, engineer new features or transform existing ones to enhance the model's predictive power. |
| Model Development | Develop a predictive model using machine learning algorithms to estimate laptop prices based on the identified features. This may involve techniques such as regression analysis, decision trees, random forests, or gradient boosting algorithms. Train the model using the prepared dataset and fine-tune its parameters for optimal performance |
| Model Evaluation and Validation | Evaluate the performance of the developed model using appropriate evaluation metrics such as mean absolute error (MAE) or root mean squared error (RMSE). Validate the model's predictive ability using cross-validation techniques and ensure its reliability and accuracy. |
| Documentation and Reporting | document the entire analysis process, including data collection, preprocessing steps, EDA findings, feature engineering techniques, model development, and evaluation results. Prepare a comprehensive report summarizing the analysis methodology, key insights, and the predictive model's performance. Present the report to the store manager and other stakeholders involved in the project. |
| Collaboration and Feedback | Collaborate with the store manager and other stakeholders throughout the project. Seek feedback and incorporate any necessary modifications or additional analyses based on their input. Regularly communicate progress, discuss findings, and provide updates on the project's status. |

## This project does not include:

*Specify the things that this project isn’t responsible for doing (out of scope). For instance, “this project does not involve a summation of 2019 data analysis”*

* the year depend on the last year will find in the sites
* just laptops any other device not included

## Deliverables:

*A specific list of things that your project will deliver.*

| Deliverable | Description/ Details |
| --- | --- |
| data collection report | A report summarizing the sources of laptop data collected, including online retailers, marketplaces, and manufacturer websites. It should provide details on the data collection methods and any challenges or limitations encountered during the process. |
| cleaned dataset | A cleaned and preprocessed dataset that ensures consistency and accuracy. This dataset will be ready for further analysis and modeling |
| EDA report | report presenting the findings from the EDA phase, including statistical summaries, visualizations, and insights into the laptop market in Algeria. This report will highlight patterns, trends, and relationships between variables that may impact laptop prices. |
| feature selection and engineering report | A report detailing the selected features and attributes that significantly influence laptop prices. It will describe the feature selection techniques used and any engineered features created to enhance the predictive model's performance. |
| predictive model | The developed predictive model, implemented using appropriate machine learning algorithms. This model will estimate laptop prices based on the identified features and will be ready for deployment or further refinement. |
| Model Evaluation and Validation report | A report evaluating the performance of the predictive model using root mean squared error (RMSE). This report will provide insights into the model's accuracy and reliability, demonstrating its predictive ability. |
| Final Project Report | A comprehensive report summarizing the entire analysis process. It will include details on data collection, preprocessing, EDA, feature selection, model development, evaluation results, and any additional analyses conducted. The report will present key insights, recommendations, and the overall predictive model's performance. |

## Schedule Overview / Major Milestones:

*The expected schedule for the project. This can be defined by milestones (e.g. “all data is cleaned and processed”), periods of time (“Week 1 / Week 2”), or other ways based on the needs of the project.*

| Milestone | Expected Completion Date | Description/Details |
| --- | --- | --- |
| *data collection* | *10/08/2023* | *completion of data collection from various sources, including online retailers, marketplaces, and manufacturer websites. It ensures that all the necessary data for analysis is acquired* |
| *Data Cleaning and Preprocessing Completed* | *20/08/2023* | *The collected data is cleaned and preprocessed, ensuring consistency and accuracy. Missing values are handled, duplicates are removed, and data formats are standardized.* |
| *Exploratory Data Analysis* | *10/09/2023* | *Statistical summaries, visualizations, and insights into the laptop market in Algeria are generated. Patterns, trends, and relationships between variables are identified.* |
| *Feature Selection and Engineering Completed* | *20/09/2023* | *the key features and attributes that significantly influence laptop prices are identified and selected. Feature engineering techniques are applied, including creating new features or transforming existing ones to enhance the predictive model's performance* |
| *Predictive Model Developed* | *03/10/2023* | *This milestone signifies the completion of developing the predictive model using machine learning algorithms. The model is trained using the prepared dataset, and its parameters are fine-tuned for optimal performance.* |
| *Model Evaluation and Validation Completed* | *20/10/2023* | *the developed predictive model is thoroughly evaluated using appropriate metrics such as mean absolute error (MAE) or root mean squared error (RMSE). Cross-validation techniques are applied to validate the model's predictive ability.* |
| *Final Project Report Submitted* | *01/11/2023* | *The final project report, summarizing the entire analysis process, including data collection, preprocessing, EDA, feature selection, model development, and evaluation results, is submitted.* |

## \*Estimated date for completion:

*This is my “if all goes well and I have everything I need, this is when I’ll be done” date.*

1/12/2023