# Project Scope

| Project | Project Manager | Date |
|---|---|---|
| **EDA** : A WebApp for Efficient Exploratory Data Analysis | Dr. Nitin Uikey | 01/10/2024 |

| Justification |
|---|
| The need for an interactive and efficient platform to conduct Exploratory Data Analysis (EDA) without coding skills has been identified. This project will provide organizations, especially in IT industries, with the ability to quickly analyze large datasets and find hidden patterns, saving time and resources. |

| Scope Description | |
|---|---|
| **In Scope** | **Must (M), Should (S), Could (C), Won't (W)** |
| **Data Uploading (CSV Format),**<br>**Basic Statistical Methods,**<br>**Data Exploration (Summarization, Visualization),**<br>**Feature Engineering (Simple),**<br>**Downloadable Graphs (.png format),**<br>**Downloadable Data Sets (.csv format)** | (M)The application will allow users to upload CSV files for analysis,<br>(M)Perform basic statistical computations such as summarization and visualization,<br>(M)Tools for visualizing data and summarizing it will be available,<br>(S)Allow for some simple transformations on data to improve its analysis,<br>(M)Users can download graphs in .png format,<br>(M)Analyzed datasets will be downloadable in CSV format |
| **Out of Scope** | **Must (M), Should (S), Could (C), Won't (W)** |
| **Advanced Statistical Components**<br>**Handling Large Datasets (>200MB)**<br>**Non-CSV File Formats**<br>**Data Cleaning for Large Datasets** | (W)Complex statistical methods, such as advanced feature engineering and handling missing data, will not be included,<br>(W)The app will not handle datasets larger than 200MB due to file size constraints,<br>(W)The application will only support CSV files for data uploads,<br>(C)This feature may be considered in future releases but is not part of the current scope. |

| Business Objectives |
|---|
| The project aims to provide organizations with the ability to perform Exploratory Data Analysis quickly, saving both time and money. By reducing the need for large teams or technical expertise in coding, the application will increase business efficiency and support better data-driven decision-making. |

| Project Deliverables |
| --- |
| **Version 1.0 of the WebApp:** Initial version including basic functionality for data uploading(CSV Format,Excel format) statistical analysis, and simple visualizations. |
| **Version 2.0 with Improved Features:** The second version will integrate faster processing, improved data visualization, and basic feature engineering tools. |
| **Real-world Testing of the WebApp:** Testing the web application on real-world datasets to ensure stability and performance. |
| **Deployment of the WebApp:** Hosting the application on the internet for public access. |
| **Documentation and User Guide:** Provide comprehensive documentation, including user instructions and troubleshooting guidelines. |

| Project Exclusions |
| --- |
| **Advanced Data Cleaning Methods:** Methods for data cleaning, particularly for larger or more complex datasets, will not be part of the initial release. |
| **Support for Non-CSV File Formats:** The web application will not support file formats other than CSV. |
| **Advanced Predictive Analytics or Machine Learning:** This version will not include predictive analytics or advanced machine learning tools. |

| Constraints |
| --- |
| **File Size Limitation:** The app can only handle datasets up to 200MB in size. |
| **Time Constraint:** The project is expected to be completed by 23/12/2024, with no significant delays in milestones. |
| **Technology Stack Limitation:** The application will be developed using a limited tech stack, with no advanced cloud infrastructure in the current version. |
| **Budget Constraints:** Limited resources for implementing additional or advanced functionalities. |

| Assumptions |
| --- |
| **Stable Internet Connection:** The web application will rely on users having a stable internet connection for optimal performance. |
| **User Knowledge of Data Formats:** Users or organizations will primarily upload structured datasets (CSV format) and will have some knowledge of how to prepare data. |
| **Limited User Base Initially:** The web application will be used by a limited number of users in the initial phase. |
| **Organizational Buy-In:** Organizations will be willing to adopt this tool for their internal data exploration and analysis processes. |

| Cost Estimate |
| --- |
| **Development Costs:** The estimated cost for developing the application, including internal labor, is approximately **$50,000**. |
| **Testing and Quality Assurance:** Costs associated with testing the web app, estimated at **$10,000**. |
| **Deployment and Maintenance:** Hosting costs and ongoing maintenance over the first 6 months are estimated at **$5,000**. |
| **Total Estimated Cost: $65,000** for the completion and deployment of the project. |