**End-to-End Process Documentation**

**Project 1: Predicting Salary Based on Years of Experience**

**Objective**

The objective of this project is to develop a predictive model that can estimate an employee's salary based on their years of experience. The purpose is to assist HR and recruitment teams in making informed decisions about salary offers and budgeting.

**Solution Architecture**

1. \*\*Data Collection:\*\* The dataset includes records of employees' years of experience and their corresponding salaries.

2. \*\*Data Profiling:\*\* Analyze the dataset to understand its structure, identify missing values, and gather statistical insights.

3. \*\*Exploratory Data Analysis (EDA):\*\* Utilize visualizations such as distribution plots and heatmaps to explore the data and understand the relationships.

4. \*\*Data Splitting:\*\* Split the data into training and testing sets to validate the model's performance.

5. \*\*Model Development:\*\* Build a Simple Linear Regression model to predict salaries based on years of experience.

6. \*\*Model Evaluation:\*\* Evaluate the model using metrics like R-squared score and residual analysis to ensure accuracy.

**Methodology**

**Data Profiling:**

\* Number of Records: 30

\* Number of Columns: 2

\* Missing Values: None

\* Data Types: Years of Experience (float), Salary (float)

**Exploratory Data Analysis:**

\* Skewness of Salary: Approximately normal distribution.

\* Skewness of Years of Experience: Slightly right-skewed.

\* Correlation between Variables: High positive correlation.

**Model Development:**

\* Linear Regression Model developed to predict salary based on experience.

\* Coefficient: Positive, indicating a direct relationship between experience and salary.

\* Intercept: Constant term that adjusts the baseline salary.

**Model Evaluation:**

\* R-squared Score: Indicates that the model explains a large portion of the variance in the data.

\* Residual Analysis: Confirms that the assumptions of linear regression hold.

**Time Taken**

The total time taken to complete this project was approximately 20 seconds, including data loading, profiling, EDA, model training, and evaluation.

**Project 2: Predicting Delivery Time Based on Sorting Time**

**Objective**

The objective of this project is to build a predictive model that estimates delivery time based on sorting time of packages. The aim is to improve operational efficiency and optimize delivery processes within the logistics department.

**Solution Architecture**

1. \*\*Data Collection:\*\* The dataset includes records of sorting times and their corresponding delivery times.

2. \*\*Data Profiling:\*\* Analyze the dataset to identify patterns, detect any missing data, and compute basic statistics.

3. \*\*Exploratory Data Analysis (EDA):\*\* Use visualizations to explore the data and uncover relationships between variables.

4. \*\*Data Splitting:\*\* Divide the data into training and testing sets to ensure robust model validation.

5. \*\*Model Development:\*\* Implement a Simple Linear Regression model to predict delivery time based on sorting time.

6. \*\*Model Evaluation:\*\* Assess the model's performance using metrics such as R-squared score and residual plots.

## Methodology

## 1. Data Profiling:

## \* Number of Records: 20

## \* Number of Columns: 2

## \* Missing Values: None

## \* Data Types: Sorting Time (float), Delivery Time (float)

## 2. Exploratory Data Analysis:

## \* Skewness of Delivery Time: Approximately normal distribution.

## \* Skewness of Sorting Time: Slightly right-skewed.

## \* Correlation between Variables: Strong positive correlation.

## 3. Model Development:

## \* Linear Regression Model developed to predict delivery time based on sorting time.

## \* Coefficient: Positive, indicating a direct relationship between sorting and delivery times.

## \* Intercept: Baseline time for delivery independent of sorting.

## 4. Model Evaluation:

## \* R-squared Score: Demonstrates the model's accuracy in explaining the variance in delivery time.

## \* Residual Analysis: Confirms that the model's assumptions are valid.

## Time Taken

## The total time taken to complete this project was approximately 18 seconds, including data loading, profiling, EDA, model training, and evaluation