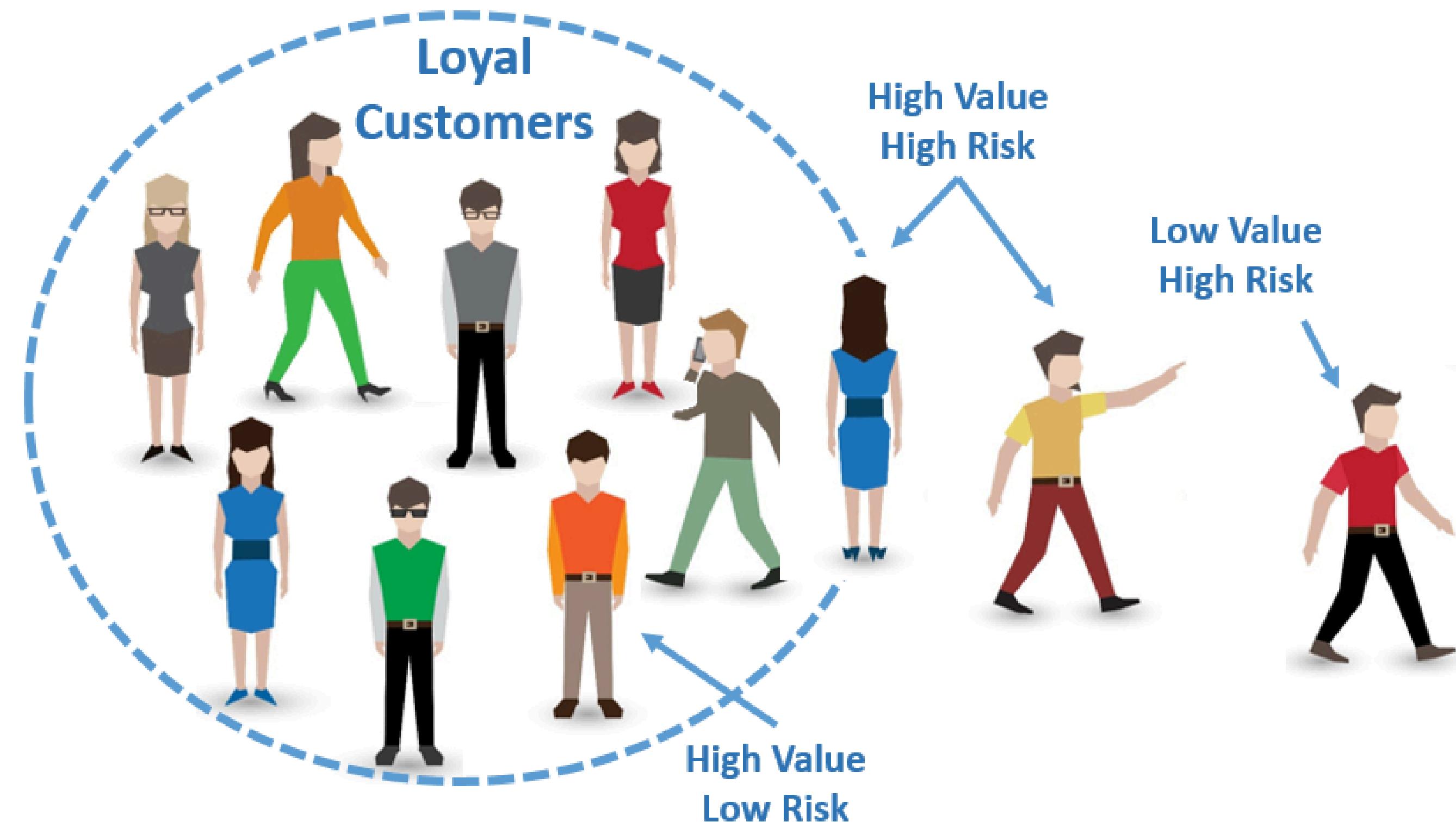
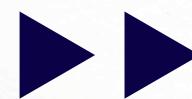


CUSTOMER CHURN PREDICTION



By: Atharva Hatekar



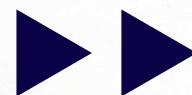
INTRODUCTION

Churn means what:



Churn rate is a critical metric of customer satisfaction. Low churn rates mean happy customers; high churn rates mean customers are leaving you.

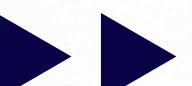
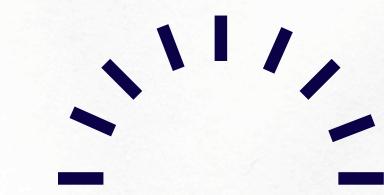
According to Google: it takes a lot more money (up to five times more) to get new customers than to keep the ones you already have. Churn tells you how many existing customers are leaving your business, so lowering churn has a big positive impact on revenue.





WORKFLOW

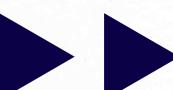
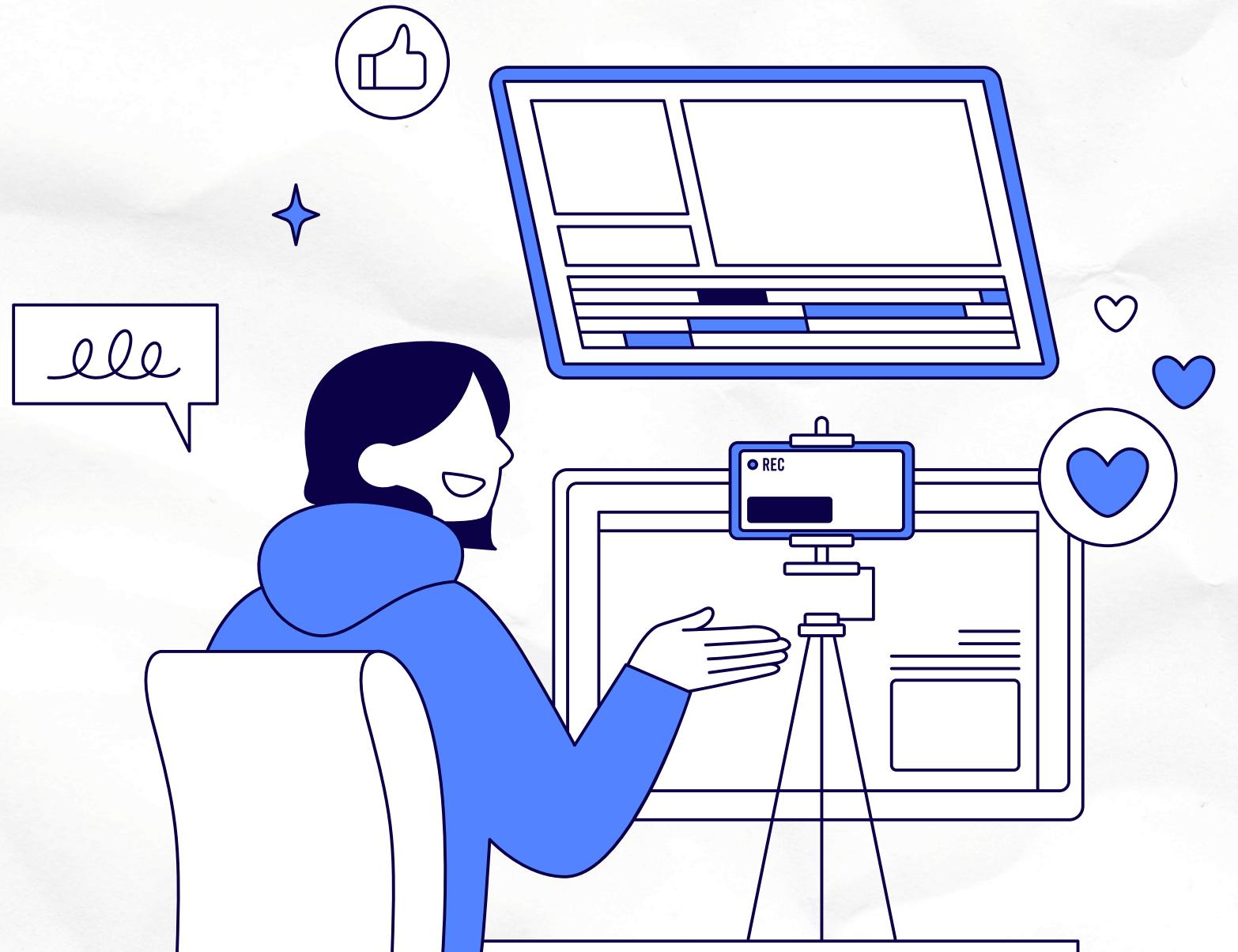
- Identify Problem Statements
- Data Collection
- Exploratory Data Analysis
- Statistical Analysis
- Feature Engineering
- ML Model Selection



TOOLS I USED



- **Python : Overall Language**
- **Pandas : Data Processing**
- **Numpy : Linear Algebra**
- **Matplotlib : Visualization**
- **Plotly : Visualization**
- **Statsmodels : Statistical analysis**
- **sklearn : Machine Learning Techniques**





ABOUT THE DATASET

- Customer churn data sourced by the IBM Developer Platform.
- It has data for 7043 Clients, with 20 features.
- Dataset consists of 17 Categorical features, 3 Numerical features and 1 target variable.
- it already Includes a target label indicating whether or not the customer left the services.

EXPLORATORY DATA ANALYSIS (EDA)



- Exploring the data to discover patterns and visualize how the features interact with the label (Churn or not).
- It Involves:
 - Checking for missing values.
 - Analyzing distributions of features.
 - Visualizing correlations.
 - Extracting key insights.



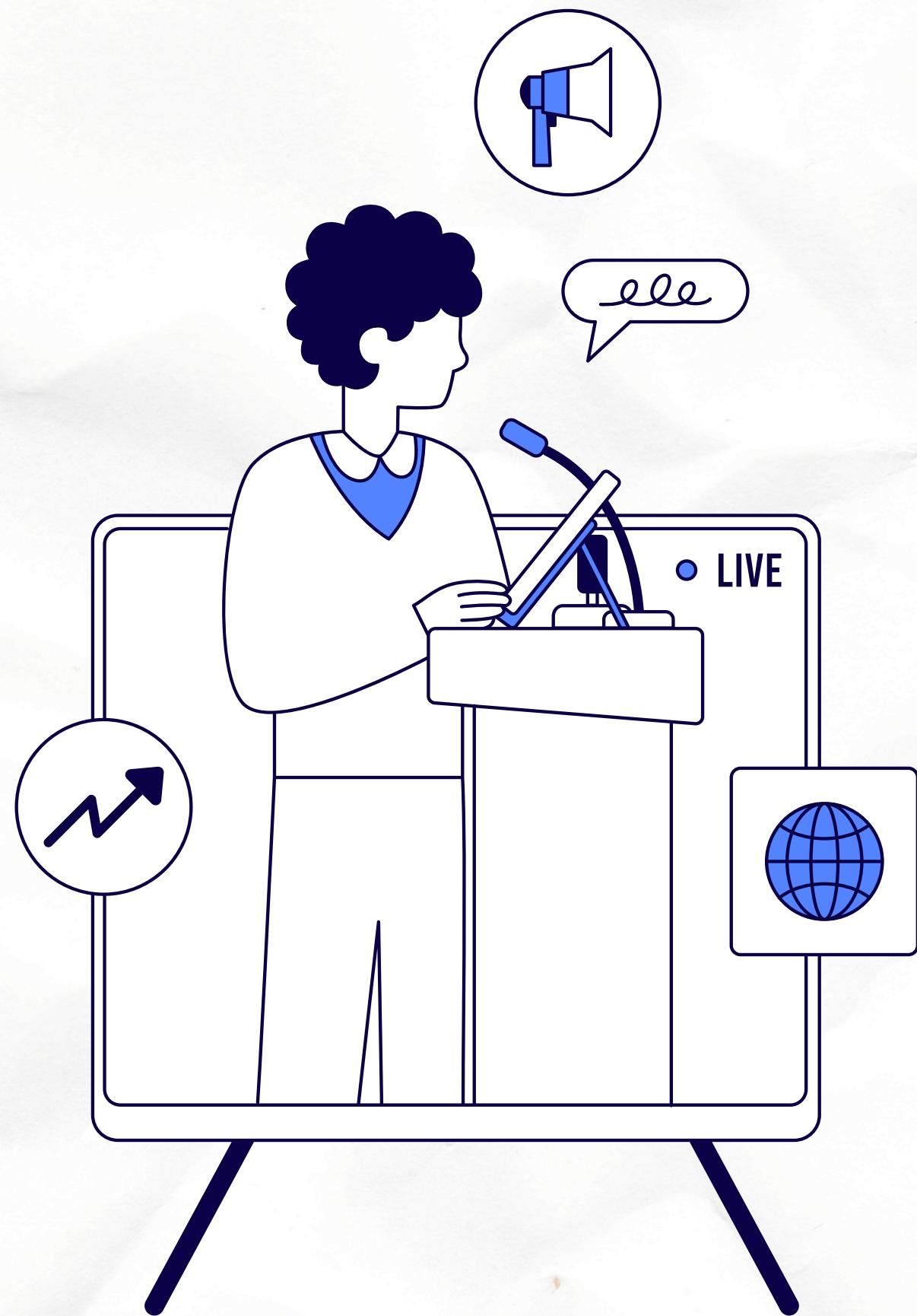
DATA PREPROCESSING



- Data from the real world is incomplete, inconsistent and generally immature and its difficult to use for Data analyses and tasks of the machine learning.
- It helps to eliminates problems such as missing values, outliers and other errors in the system. When done correctly, data preprocessing lays the foundation for accurate results with meaningful insights.

STATISTICAL ANALYSIS

- finding correlations between features to reveal patterns and trends.
- Features with high correlation are more linearly dependent and have almost the same effect on the dependent variable. So, when two features have a high correlation, dropping one of them can help to increase the efficiency of ML-Model.



MACHINE LEARNING MODELS



- Evaluating different models helps to determine the best Machine Learning model for a corresponding dataset
- The reason we have multiple metrics and ways to compare the models is that for different problems, different metrics will be more relevant.
- Choosing a right ML Model is a hit and try method.



THANK
YOU