Proposal

Question/need:

Missing appointment happens a lot and sometimes affect businesses that depend on appointments to complete the service such as Hospitals. In this project we will build a model based on our dataset to predict the probability of the patient to miss the appointment.

The benefit of this model that the hospital could predict the probability of the patient to miss the appointment and then increase number of appointment for that day.

Data Description:

The dataset is provided by **Kaggle** and contains more than 10000+ records with 14 features here are some of important features:

- **Gender**
- **AppointmentDay**
- **Neighborhood**
- **Scholarship**
- **Hypertension**
- **Diabetes**
- **Alcoholism**
- **Handcap**
- **SMS_received**
- **No-show**

we will see what is the relation between these features and NO_SHOW

Dataset url: https://d17h27t6h515a5.cloudfront.net/topher/2017/October/59dd2e9a_noshowappointments-kagglev2-may-2016/noshowappointments-kagglev2-may-2016.csv

Tools:
- Jupyter notebook
- matplotlib
- Pandas
- Numpy
- sklearn

MVP Goal:

- Load Libraries : Import all needed libraries for the project
- Data Wrangling : Load the file and check if there any null values or mistakes in the data
- Data Cleaning : Removing unwanted features and clean the data
- Data Exploratory : Analyze and explore the relationship between the features
- Data Modeling : Generate and evaluate and test the model
- Data Limitation : Discuss the limitation of the model and the dataset
- Conclusion : Discuss the findings