Heart Disease Prediction Project Documentation

1. Project Planning

Project Objective:

The project aims to analyze heart disease data and predict whether an individual has heart disease based on health and demographic indicators such as age, gender, blood pressure, cholesterol level, heart rate, and lifestyle habits.

Tools and Technologies:

- Power BI: For data analysis and interactive dashboard visualization of historical patient data.
- AI Model: A classification model predicting Yes / No for heart disease on new user data.
- Power Apps: For allowing users to input personal data and receive a prediction along with the main contributing factors.

Project Phases:

1. **Data Preparation:**
    - Cleaning the dataset.
    - Removing outliers and invalid values.
    - Ensuring data accuracy.
2. **Exploratory Data Analysis (EDA):**
    - Studying relationships between health variables.
    - Identifying key risk factors affecting heart disease.
3. **Model Building:**
    - Developing a classification model to predict whether an individual has heart disease (Yes / No).
    - Evaluating model performance using metrics such as accuracy, precision, and recall.
4. **Dashboard Development:**
    - Analyzing historical patient data from the dataset.
    - Presenting insights through interactive visualizations (charts, graphs, and tables).
5. **Integration with Power Apps:**
    - Users input their personal health data.
    - The system predicts whether they have heart disease or not using the AI model.

2. Stakeholder Analysis

Problem Statement:

Heart disease is one of the leading causes of death worldwide. It often results from multiple interacting factors such as high cholesterol, high blood pressure, aging, and unhealthy lifestyle habits. There is a lack of precise understanding of the impact of these factors on individual risk.

Project Goal:

- Identify the most influential factors associated with heart disease (e.g., age, cholesterol, blood pressure) from historical data.
- Build a predictive model to classify individuals as having or not having heart disease.

Stakeholders:

- Doctors and Health Authorities: Use data-driven insights for early prevention and better patient care.
- Individuals/Patients: Understand personal risk and take proactive steps to improve health.
- Data Analysts/Researchers: Analyze trends and risk factors for further studies.

Expected Impact:

- Support early prevention strategies.
- Provide personalized health predictions.
- Enable informed decision-making in healthcare.

3. Database Design

Dataset Source:

- The dataset is publicly available on Kaggle.
- It contains data on patients with various health indicators related to heart disease.

Columns:

Column Name	*Type*	*Description*
Age	INT	Patient age
Sex	VARCHAR(10)	M / F
ChestPainType	VARCHAR(10)	ATA / ASY / NAP / TA
RestingBP	INT	Resting blood pressure
Cholesterol	INT	Cholesterol level
FastingBS	INT	Fasting blood sugar
RestingECG	VARCHAR(20)	ECG type
MaxHR	INT	Maximum heart rate
ExerciseAngina	VARCHAR(5)	Y / N
Oldpeak	FLOAT	ST depression
ST_Slope	VARCHAR(10)	Up / Down / Flat
HeartDisease	INT	Target variable (0 = No, 1 = Yes)

Design Notes:

- HeartDisease is the target variable for classification.
- Data types represent numeric, categorical, and binary variables.

4. UI/UX Design

Page 1: Summary and Key Performance Indicators (KPIs)

• Calculate Overall Prevalence Rate: Percentage of patients with heart disease (HeartDisease = 1) out of the total sample.

• Analyze Prevalence by Sex: Breakdown of the heart disease rate (HeartDisease) by the (Sex) category.

• Identify Critical Averages: Calculate the average Age, MaxHR, and RestingBP for the diseased group (HeartDisease = 1).

• Behavioral Risk Distribution: Display the proportion of patients who experience exercise-induced angina (ExerciseAngina = Y).

Page 2: Demographic and Clinical Characteristics Analysis

• Analyze Age Impact: Segment patients into age groups and calculate the heart disease rate (HeartDisease) for each group.

• Compare Chest Pain Types: Analyze the distribution of ChestPainType and compare it between diseased and healthy individuals.

• Compare Vital Metrics: Compare the averages of (RestingBP, Cholesterol) between the diseased and non-diseased groups to identify clear differences.

• Determine Optimal Heart Rate Range: Analyze the relationship between MaxHR and the probability of developing heart disease (HeartDisease).

Page 3: Advanced Clinical Factors Analysis

• Analyze ST_Slope Gradient: Analyze the relationship between Oldpeak and ST_Slope (stress test indicators) and their effect on HeartDisease.

• Effect of Fasting Sugar on Cholesterol: Compare the distribution of Cholesterol levels based on the FastingBS status (0 or 1).

• Analyze Resting ECG Status: Distribute HeartDisease cases based on the RestingECG result.

• Correlation Matrix: Calculate and display the Correlation Matrix between all numerical variables to identify the strongest factors associated with the disease.

Page 4: Risk Analysis and Simulation

• Build a Predictive Risk Model: Train a machine learning model (e.g., Logistic Regression) to predict the probability of developing heart disease (HeartDiseas).

•Enable Interactive Simulation: Allow the user to input values for specific variables (e.g., $Age, RestingBP, ExerciseAngina$) and display the predicted result (risk percentage).

• Slice Analysis: Provide tools to filter and analyze the prevalence rate for custom-defined subgroups based on multiple factors simultaneously.

---

5. Dashboard Visualization

Purpose:

The dashboard analyzes historical patient data from 918 patients to reveal key trends, correlations, and heart disease risk factors. It provides **insights into demographics, health metrics, and diagnostic indicators** that contribute to heart disease prevalence.

1. Key Metrics & Overview

Metric	*Value*	*Insight*
Total Patients	918	Sample size of the analysis
Heart Disease %	55.34%	Over half of patients are affected
Heart Disease Patients	508	Count of patients with heart disease
Healthy Patients	410	Count of patients without heart disease
Average Cholesterol	243.20 mg/dL	Above healthy range (<200 mg/dL), major risk factor

2. Heart Disease by Age Group

Age Group	*Heart Disease %*	*Insight*
Young Adults	32.50%	Lowest risk group, still significant
Middle Age	50.77%	Moderate prevalence
Older Adults	73.12%	Highest risk; age strongly correlates with heart disease

Data Visualization:

- **Bar chart** showing % of heart disease by age group.
- **Insight:** Risk increases with age.

3. Heart Disease by Sex & Age Group

- **Males:** Higher prevalence in all age groups.
- **Females:** Lower prevalence but risk increases with age.
- **Highest Risk:** Older Adult Males.

Data Visualization:

- **Grouped bar chart** (Sex × Age Group × Heart Disease %).
- **Insight:** Males are generally at higher risk than females.

4. Heart Disease by Chest Pain Type

Chest Pain Type	*Insight*
ASY (Asymptomatic)	Strongly associated with heart disease
NAP (Non-Anginal Pain)	Mixed distribution
ATA & TA	Higher proportion of healthy patients

Data Visualization:

- **Stacked bar chart** showing heart disease vs chest pain type.
- **Insight:** ASY is a major red flag.

5. Heart Disease vs Fasting Blood Sugar

- Patients with heart disease more likely to have **high fasting blood sugar (>120 mg/dL)**.
- Contribution is noticeable but secondary to age, sex, cholesterol, and ECG indicators.

Data Visualization:

- **Grouped bar chart** for high vs normal FBS by heart disease status.

6. Heart Disease by ST_Slope (ECG Indicator)

- **Flat / Down Sloping ST_Slope:** Most patients have heart disease → **strong correlation**.
- **Up Sloping:** Majority are healthy → considered safe.

Data Visualization:

- **Stacked bar chart** for Heart Disease vs ST_Slope.
- **Insight:** ST_Slope is a powerful predictor.

7. Average Resting Blood Pressure by Age Group

- Blood pressure is high across all groups, increasing with age.
- Older Adults have the highest average RestingBP → contributes to high heart disease prevalence.

Data Visualization:

- **Bar chart** for average RestingBP by age group.

8. Summary of Insights

1. **High Prevalence:** 55% of patients have heart disease.
2. **Critical Risk Factors:** High cholesterol and high blood pressure are widespread.
3. **Demographic Risk:** Older Adult Males are the most vulnerable.
4. **Strong Predictors:** ASY chest pain and Flat/Down ST_Slope on ECG.
5. **Co-morbidities:** High fasting blood sugar and hypertension are common.
