Predicting Heart Disease with Machine Learning

-Jasvanth. S

Project Goal

Why are we here?

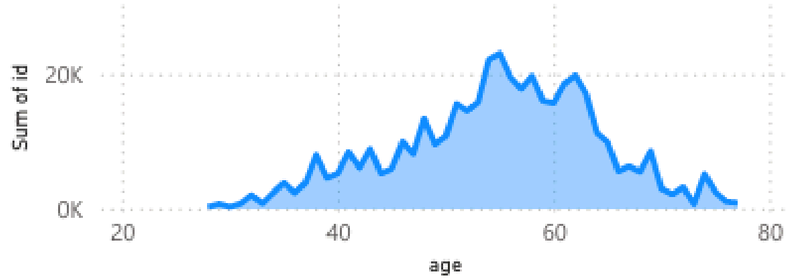
- **Objective:** Build a predictive model for heart disease.
- **Model:** Random Forest Classifier to identify patterns in patient data.
- **Impact:** Helps pinpoint key risk factors and enables early detection.

The Dataset

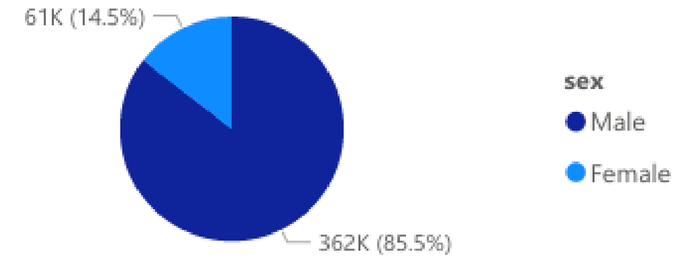
Heading: What did we use?

- **Dataset:** Heart Disease Dataset (Kaggle).
- **Features:** Age, cholesterol, chest pain type (cp), number of major vessels (ca), and more.
- **Target Variable:** Presence or absence of heart disease.

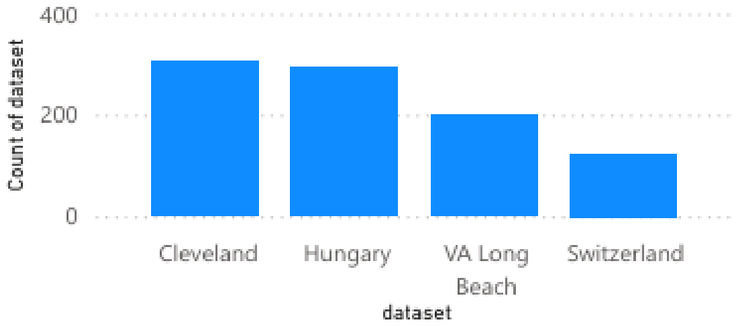
Sum of id by age



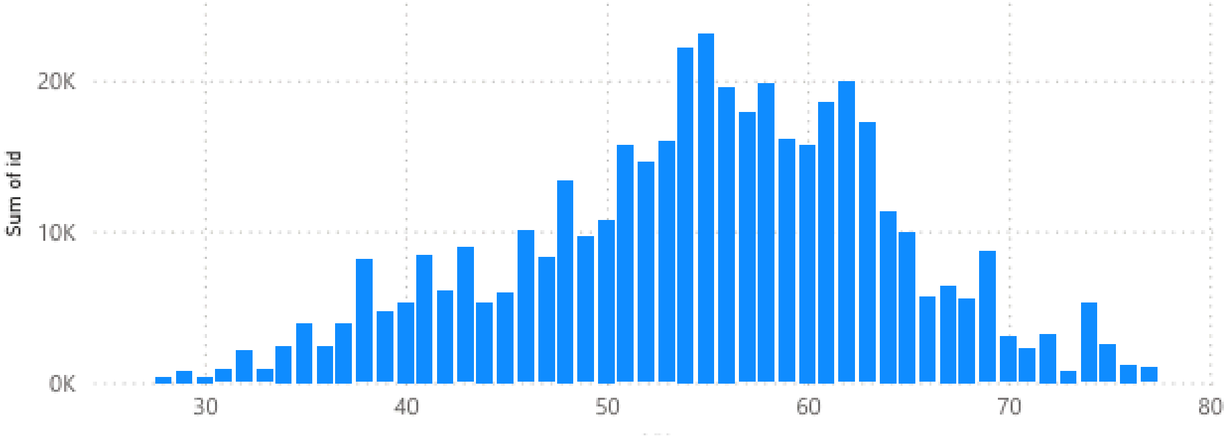
Sum of id by sex



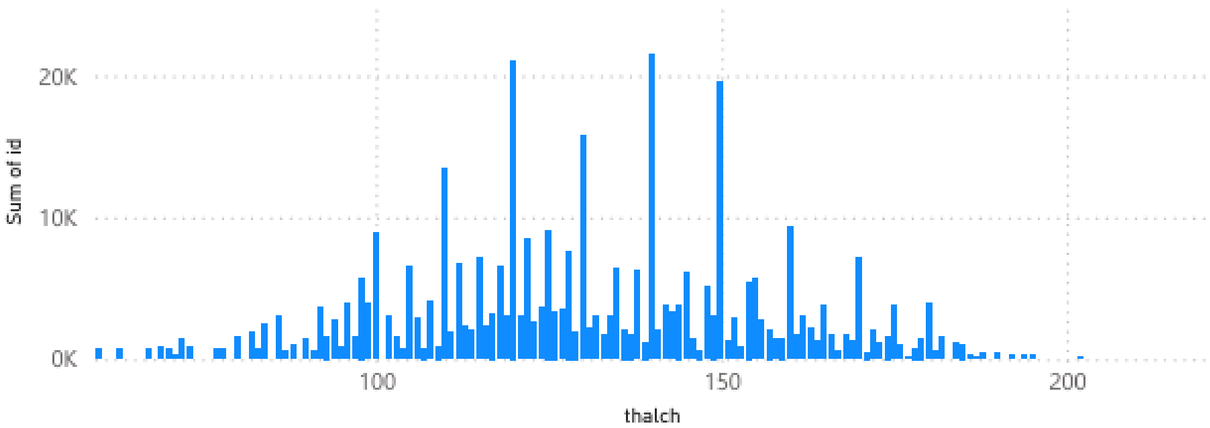
Count of dataset by dataset



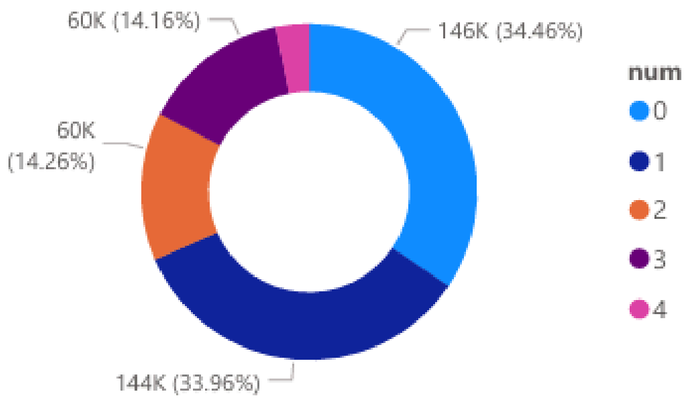
Sum of id by age



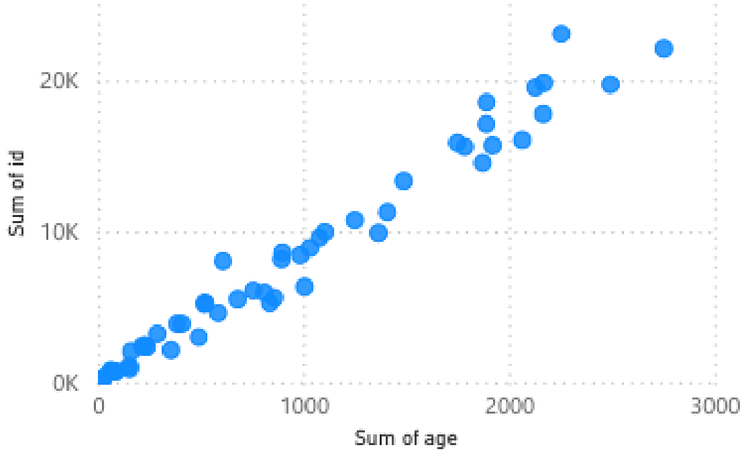
Sum of id by thalch



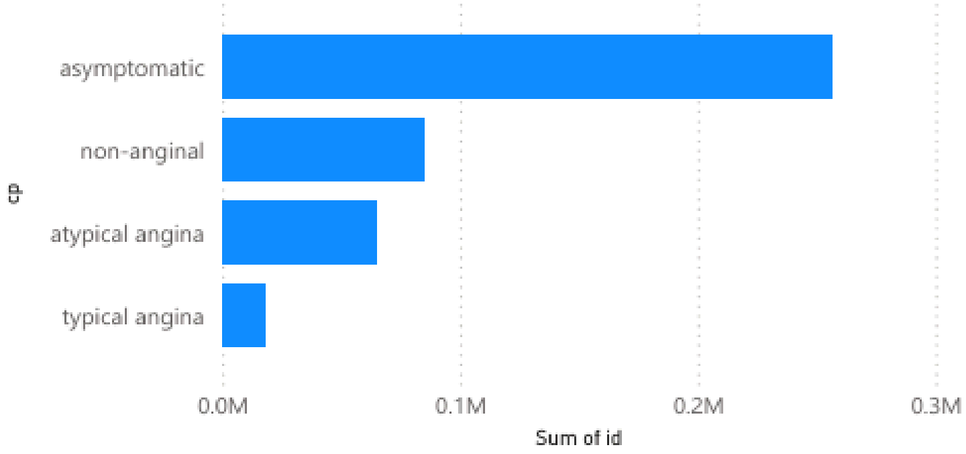
Sum of id by num



Sum of age and Sum of id by age



Sum of id by cp



Our Approach

The Machine Learning Pipeline

Data Preparation

- Handled missing values.
- Encoded categorical labels.
- Scaled numerical data.

Training

- Train-test split: 80/20.
- Trained Random Forest Classifier.

Evaluation

Checked model accuracy & performance metrics.

Model Performance

How did the model do?

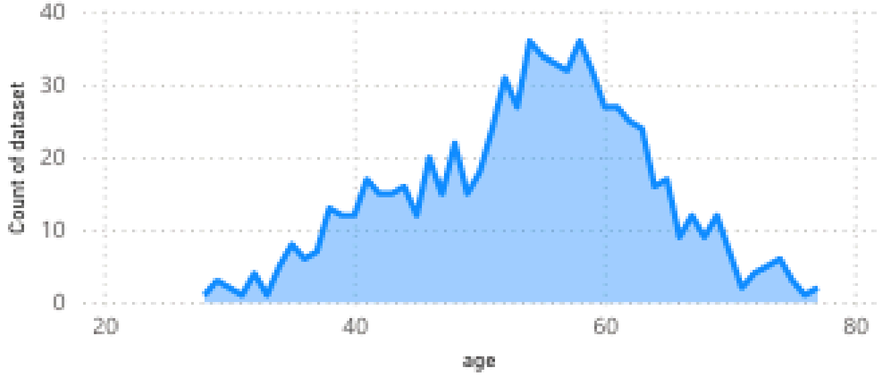
- **Accuracy:** ~88.6%.
- **Metrics:** Strong performance on both classes (disease vs no disease).
- **Takeaway:** Random Forest is a reliable predictive model.

What Matters Most?

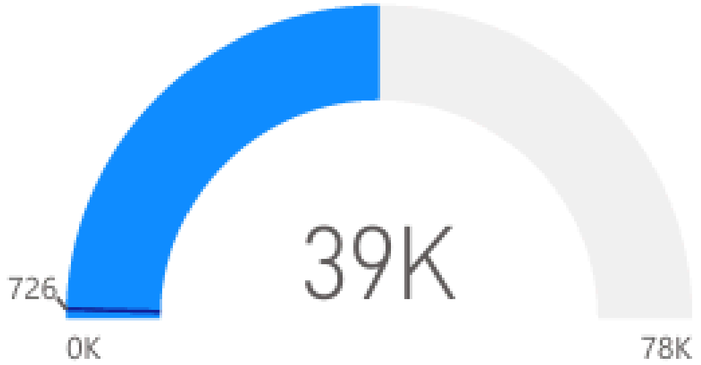
Feature Importance

- Not all features contribute equally.
- **Top Predictors:**
  - ca (major vessels)
  - thal (blood disorder)
  - oldpeak (ECG-related condition)
  - cp (chest pain type)
- **Insight:** These drive the model’s predictions.

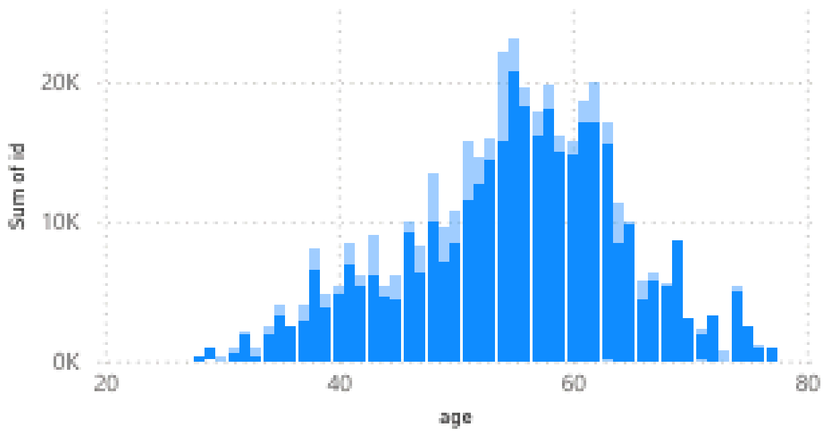
Count of dataset by age



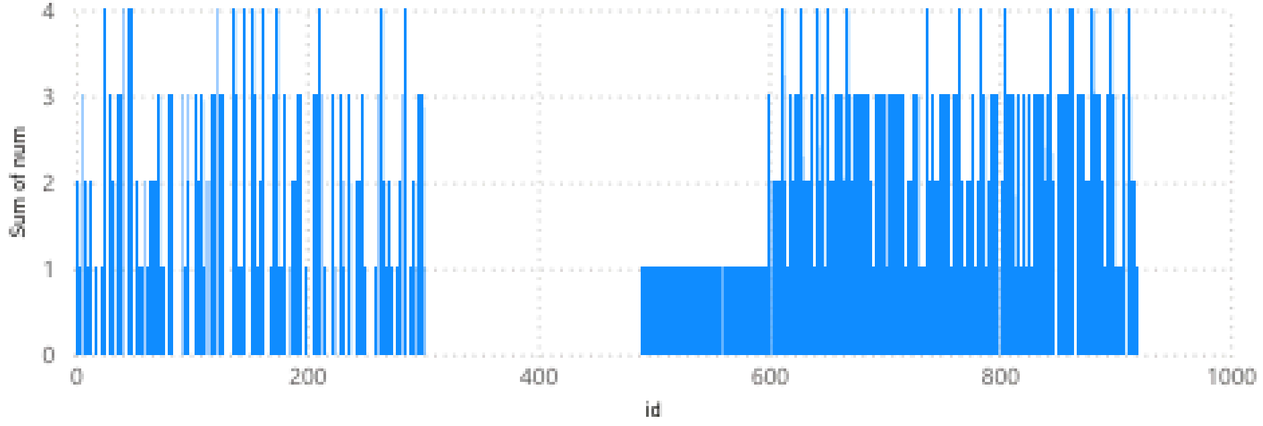
Sum of age and Count of sex



Sum of id by age

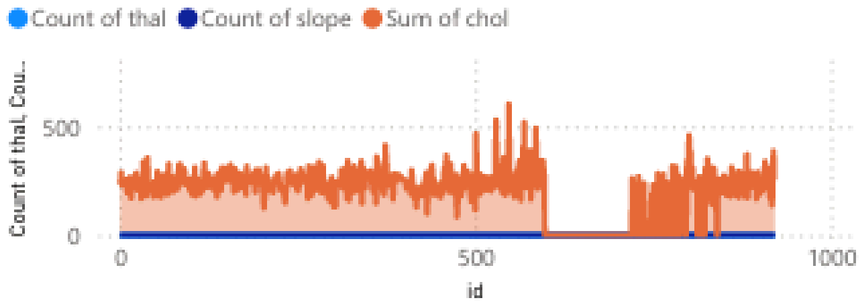


Sum of num by id

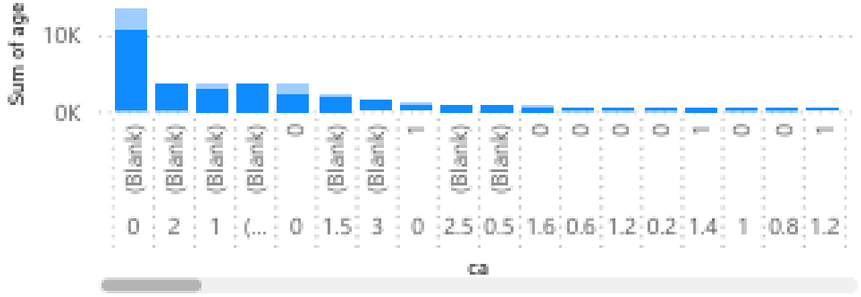


Sum of age by sex

Count of thal, Count of slope and Sum of chol by id



Sum of age by oldpeak and ca



362K

Sum of id

Next Steps :

Where do we go from here?

- **Deployment:** Save model & scaler for real-world use.
- **Prediction:** Demonstrated loading the model for new patient data.
- **Future Ideas:**
  - Experiment with other models (XGBoost, Neural Nets).
  - Train on larger datasets for improved reliability.

Thank You.

jasvanth1010@gmail.com