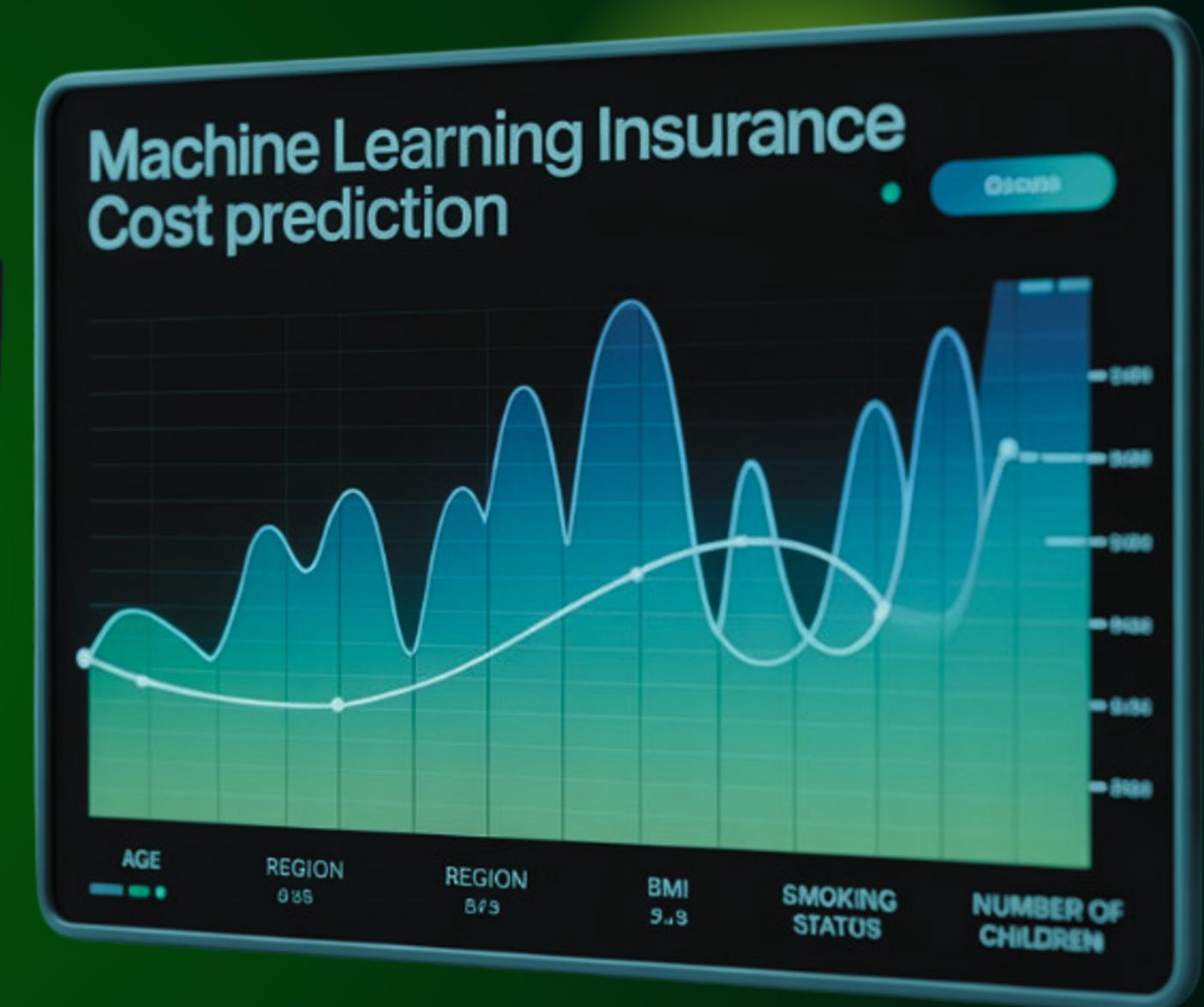


# PREDICTING AND ANALYZING HEALTH INSURANCE CHARGES USING ML & POWER BI

A CAPSTONE PROJECT USING PYTHON,  
EXCEL AND POWER BI

BY - ARNAB BANDYOPADHYAY



# PROJECT OVERVIEW

- PREDICT HEALTH INSURANCE CHARGES BASED ON DEMOGRAPHIC & LIFESTYLE FEATURES
- USE OF EDA, ML MODELING, AND POWER BI DASHBOARDING
- HELPS INSURANCE FIRMS OPTIMIZE PRICING AND UNDERSTAND RISK SEGMENTS

# PROBLEM STATEMENT

- DIFFICULTY IN FORECASTING INDIVIDUAL  
HEALTHCARE COSTS
- NEED FOR PERSONALIZED PREMIUM  
ESTIMATION
- OBJECTIVE: PREDICT CHARGES (INR) USING 7  
FEATURES (AGE, BMI, SMOKER, ETC.)



# EXPLORATORY DATA ANALYSIS (EDA)

-  **SMOKERS PAY UP TO 3X MORE THAN NON-**

**SMOKERS.**

-  **OBESE INDIVIDUALS FACE HIGHER INSURANCE CHARGES.**

-  **SOUTHEAST & NORTHEAST REGIONS SHOW**

**SLIGHTLY HIGHER AVERAGE COSTS.**

-  **AGE SIGNIFICANTLY IMPACTS CHARGES, ESPECIALLY FOR SMOKERS AND OBESE CUSTOMERS.**

Smoker

yes

no



0K

5K

Average of charges in INR

324 (24.22%)

364 (27.2%)

325 (24.29%)

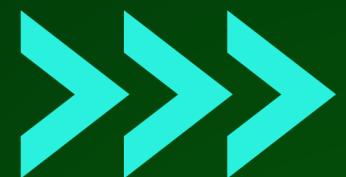
325 (24.29%)

Region

- southeast
- northwest
- southwest
- northeast

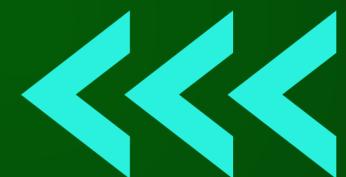
# MACHINE LEARNING WORKFLOW

DATA CLEANING



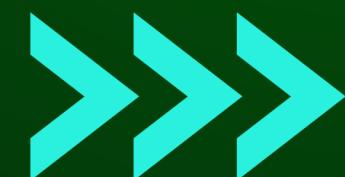
FEATURE ENGINEERING

TRAIN-TEST SPLIT



PREPROCESSING

MODEL SELECTION



HYPERPARAMETER TUNING

# MODEL EVALUATION METRICS

METRIC

VALUE

MAE(MEAN ABSOLUTE ERROR)

2530.88

RMSE (ROOT MEAN SQUARED ERROR)

4501.17

R<sup>2</sup> SCORE (COEFFICIENT OF DETERMINATION)

0.8695

TRAINING

# POWER BI DASHBOARD

 **TOTAL POLICIES**

**1338**

 **AVERAGE CHARGES**

**13.27K**

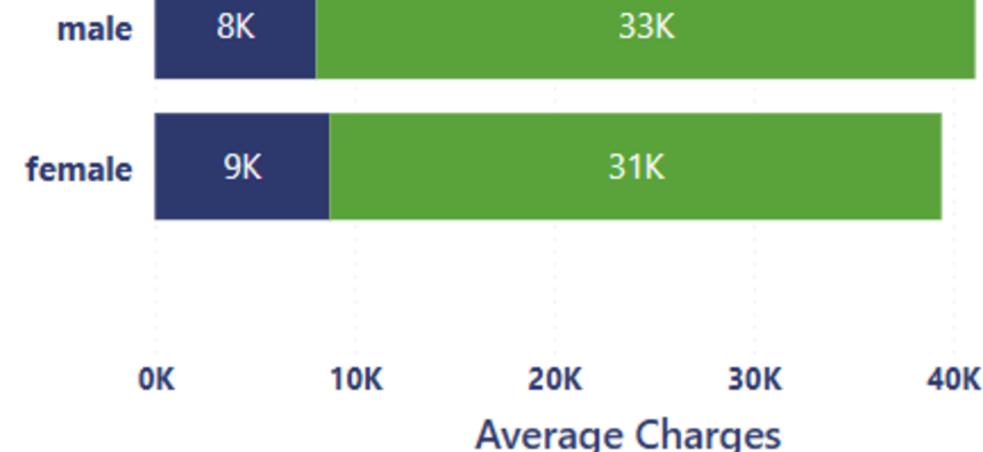
 **AVERAGE BMI**

**30.66**

## Health Insurance Overview – Demographics, Charges & Risk

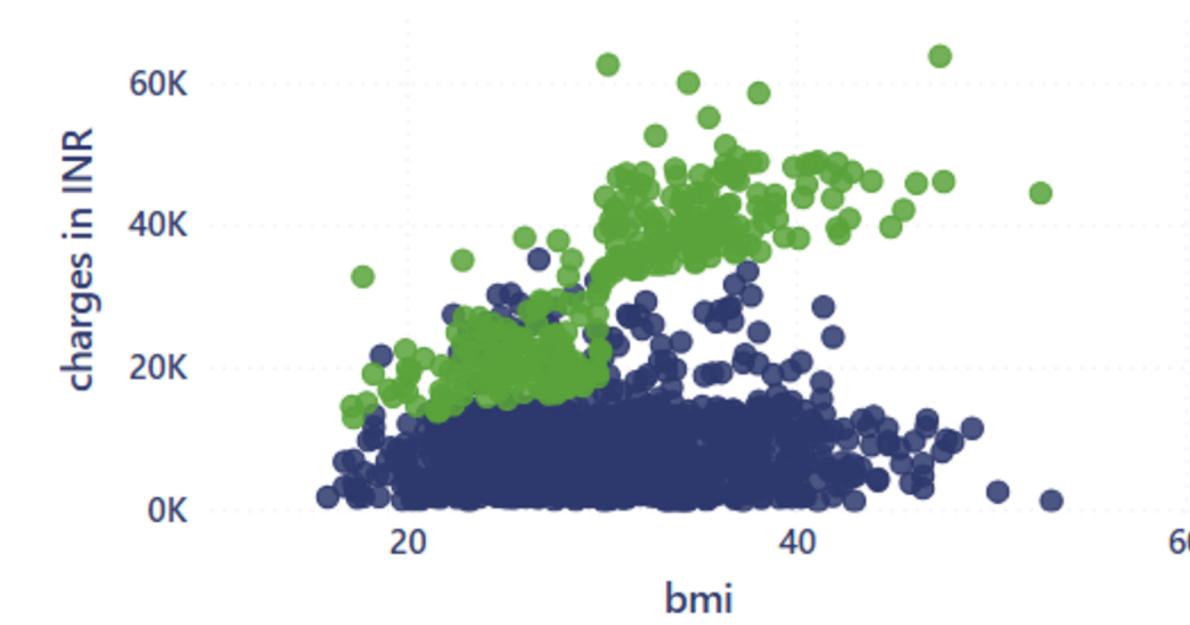
### Average Charges by gender and Smoker

Smoker • no ● yes

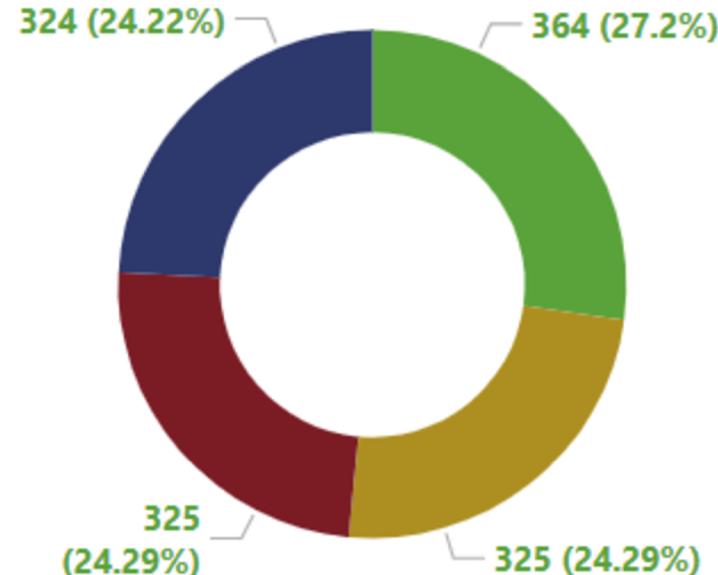


### BMI vs Charges (by Smoker)

Smoker • no ● yes

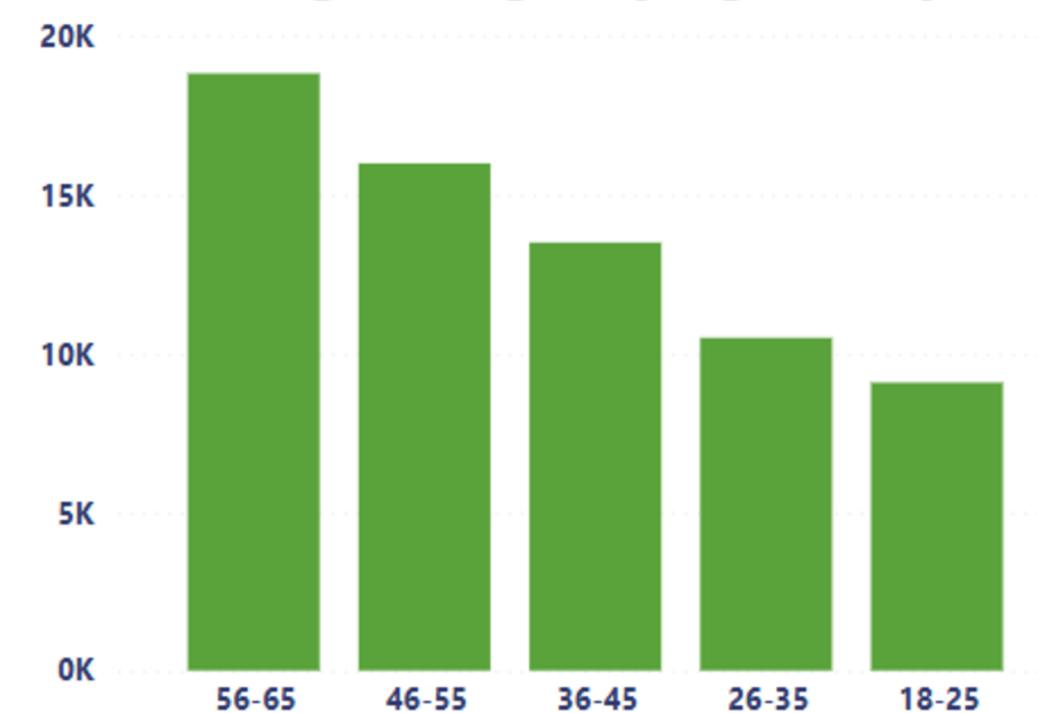


### Total Policies by Region



### Average Charges by Age Group

Region  
● southeast  
● northwest  
● southwest  
● northeast



 **Smokers (%)**

**20.48%**

 **Obesity Rate (%)**

**53.74%**

# TOOLS AND TECH STACKS

<u>CATEGORY</u>	<u>TOOLS</u>	<u>PURPOSE</u>
Programming Language	Python (pandas, sklearn, joblib)	Data processing, modeling
Spreadsheet Tool	Excel	EDA, Pivot analysis
BI Tool	Power BI	Dashboard & report visualization
IDE	Jupyter Notebook	Model building, experimentation
Version Control	GitHub	Repo hosting, version tracking

# FUTURE SCOPE

- Deploy as Web App using Streamlit or Flask for real-time user interaction
- Experiment with Advanced Models like XGBoost, SVR, or ensemble techniques
- Enhance Dataset with features like claim history, income level, and health metrics
- Scale to Real-World Data from actual insurance databases for broader impact
- Integrate with BI Tools for automated business reporting and alerts

# LET'S CONNECT

✉️ Email: [arnab.bandyopadhyay01@gmail.com](mailto:arnab.bandyopadhyay01@gmail.com)

🔗 LinkedIn: [linkedin.com/in/arnab-bandyopadhyay](https://linkedin.com/in/arnab-bandyopadhyay)

💼 About Me: [arnab-portfolio](#)

💻 GitHub: [github.com/ArnabRupam](https://github.com/ArnabRupam)

