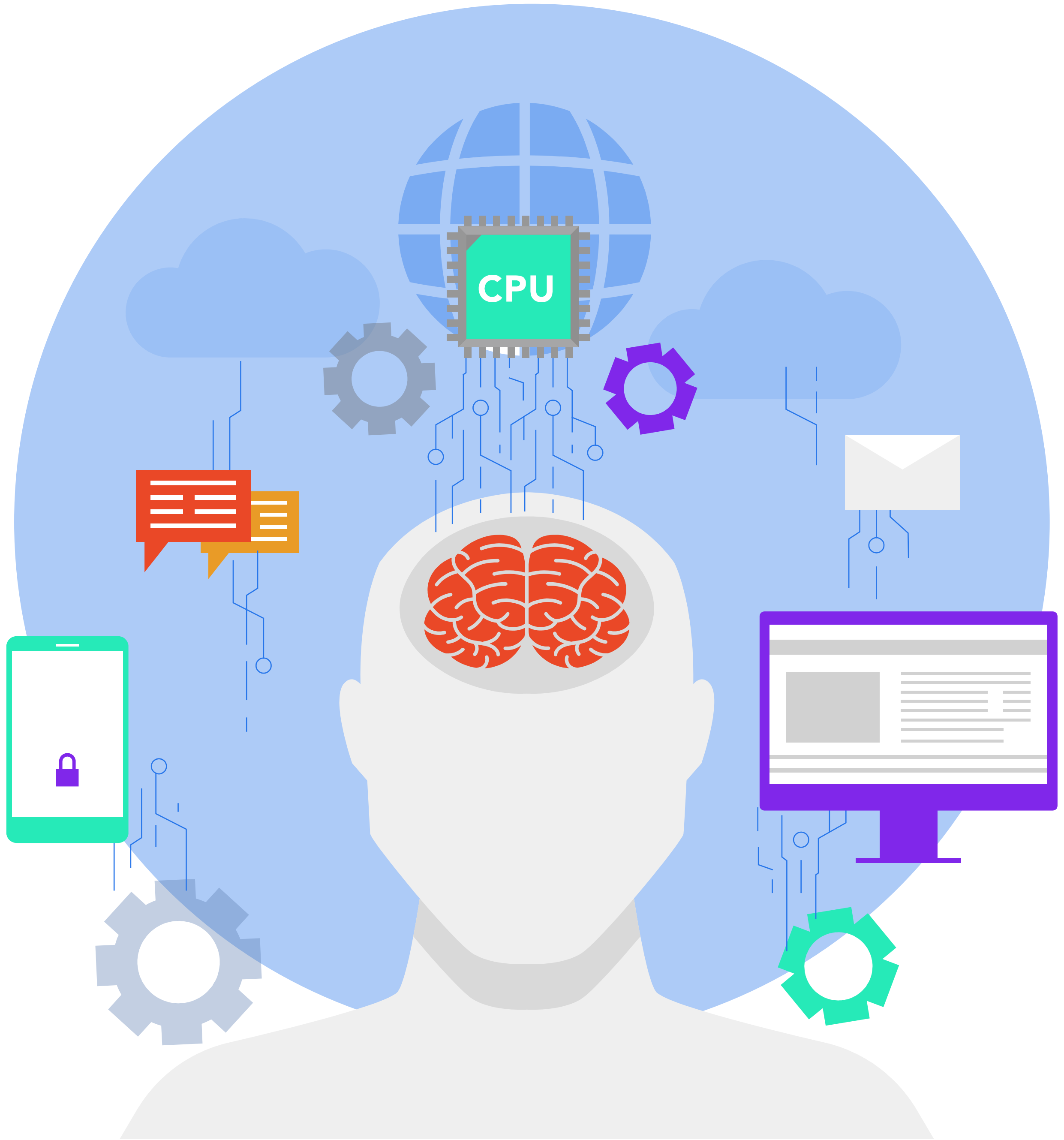


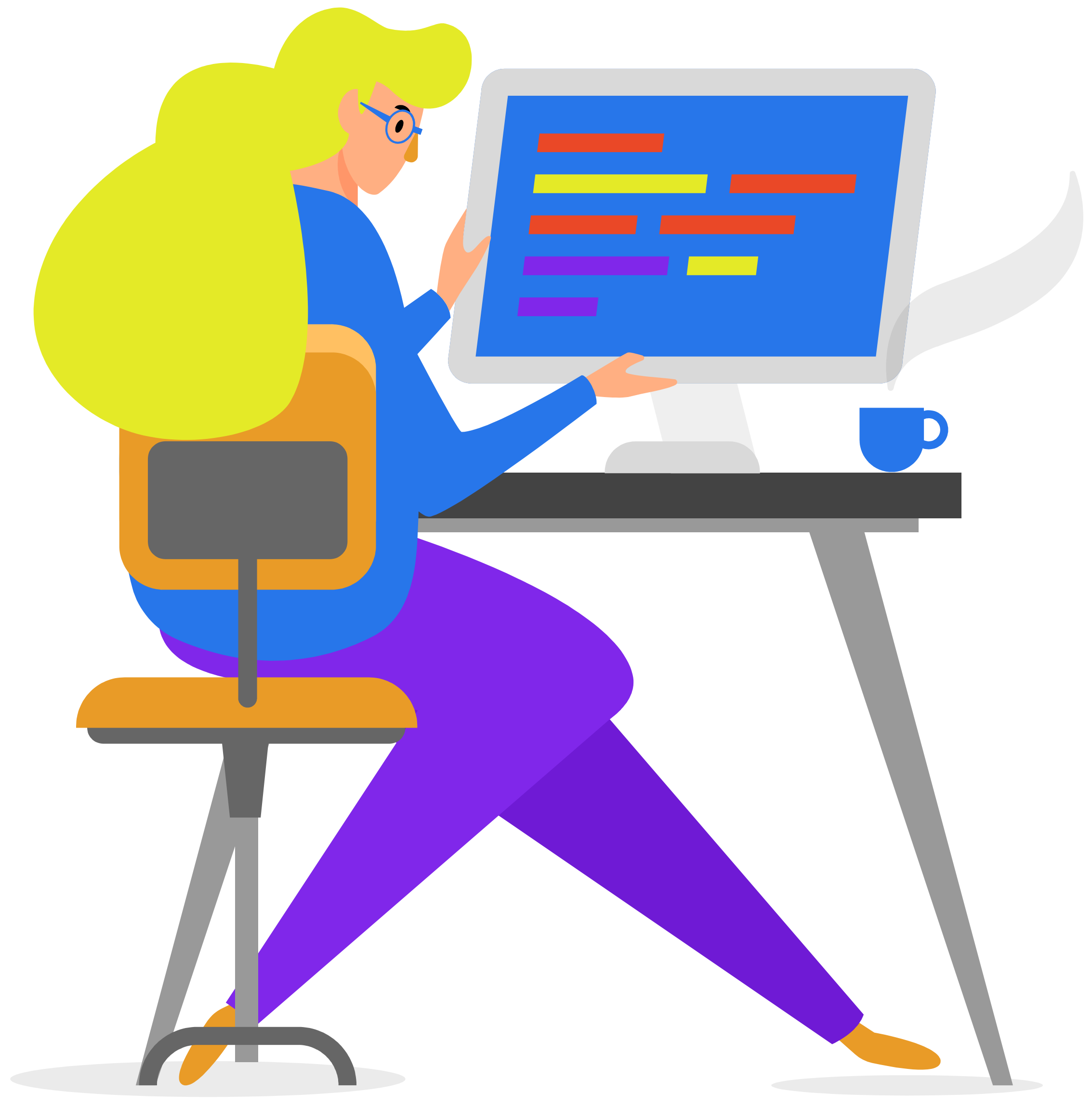
Insurance Premium Prediction

Under Guidence
Dr.Shantanu Pathak
Mrs.Priti Mam
Mrs. Priyanka Mam

Presented By
Shraddha Rananavare
Aradhya Srivastava

Group Id : 24

Contents



01 Introduction

02 Workflow of Project

03 Dataset Details

04 EDA

05 Time Series challenges & solutions

06 Model Selections & Training

07 Results & Metrix

08 Model Comparison

09 UI & Deployment

10 Requirements & Specifications

11 Outcomes

12 Conclision

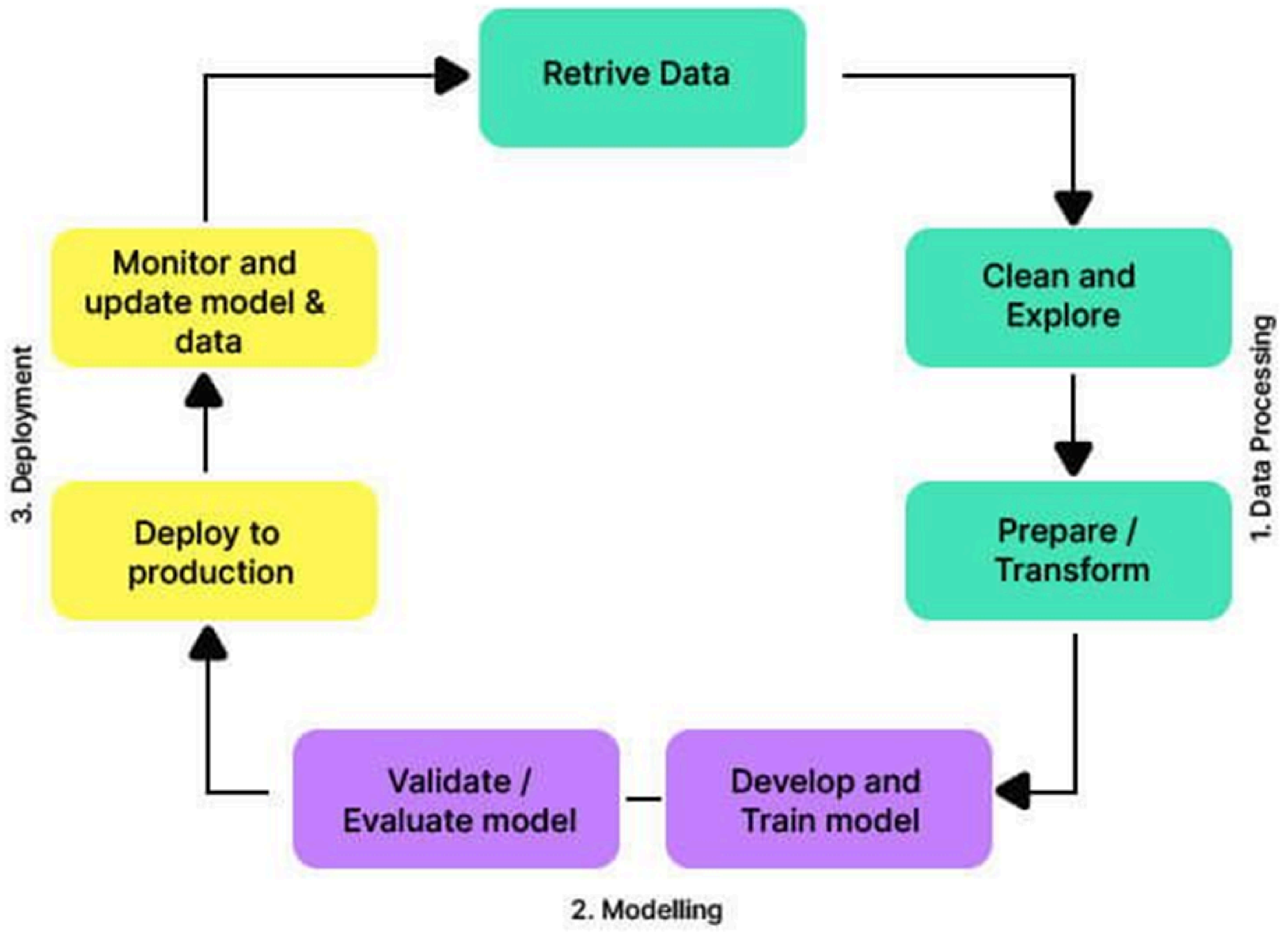
Introduction

Why Predict Insurance Premiums?

- ML models help predict accurate premiums, improving risk assessment and pricing.

By analysing features, the model will be able to estimate the price of insurance premiums for new customers.

Workflow of Project



Dataset Details

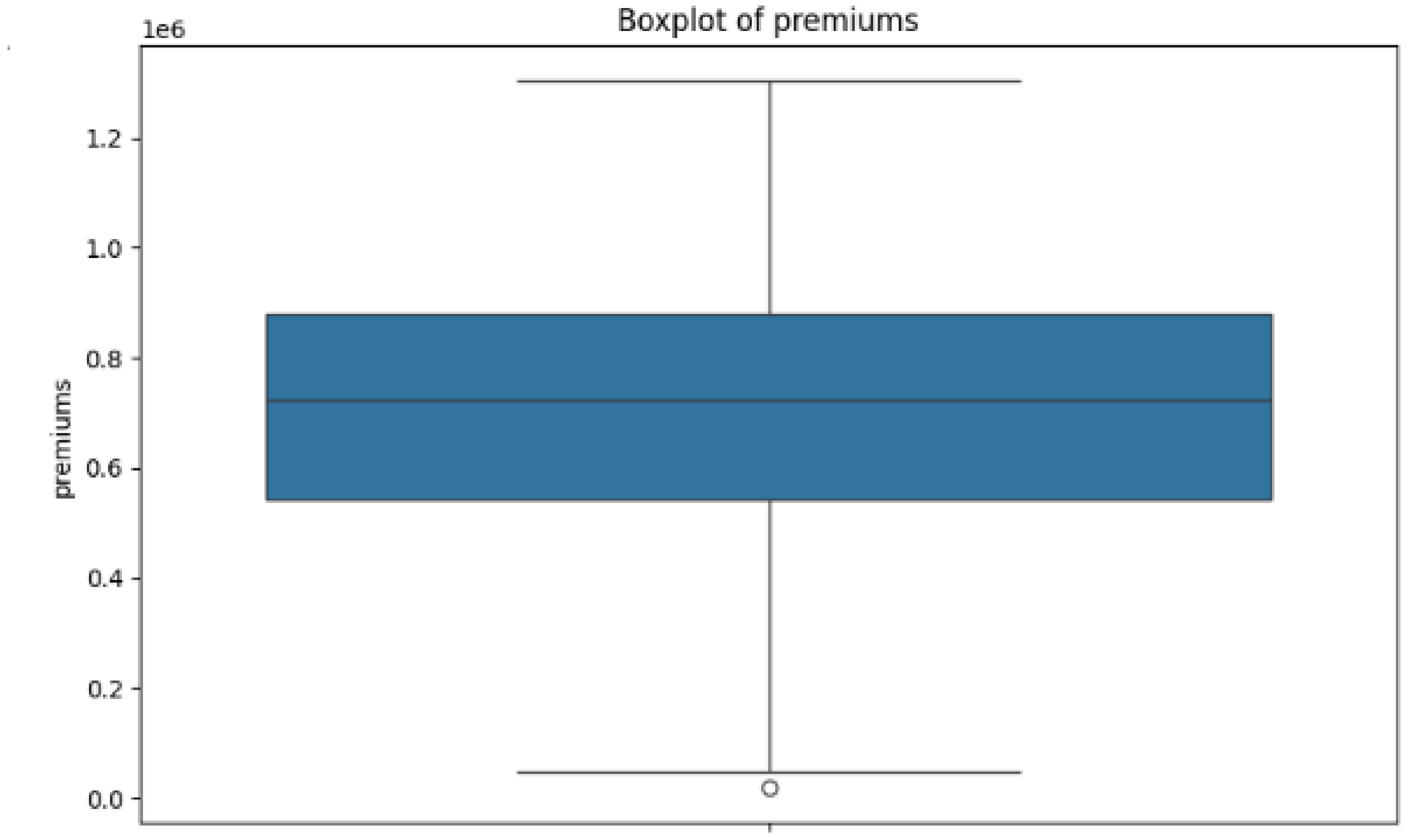
```
df.info()
```

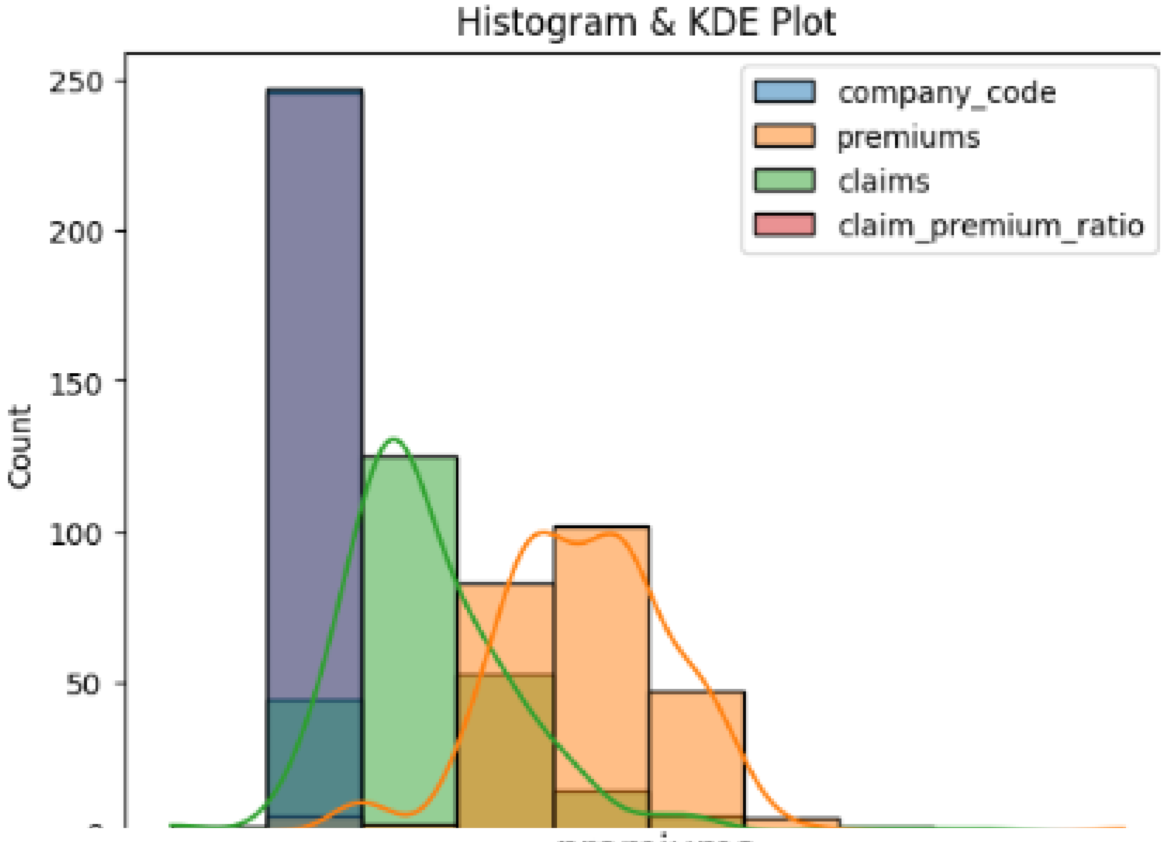
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8338214 entries, 0 to 8338213
Data columns (total 8 columns):
 #   Column               Dtype
---  ------               -----
 0   company_code         int64
 1   company_name         object
 2   year_month           object
 3   product              object
 4   state                object
 5   premiums             float64
 6   claims               float64
 7   claim_premium_ratio  float64
dtypes: float64(3), int64(1), object(4)
memory usage: 508.9+ MB
```

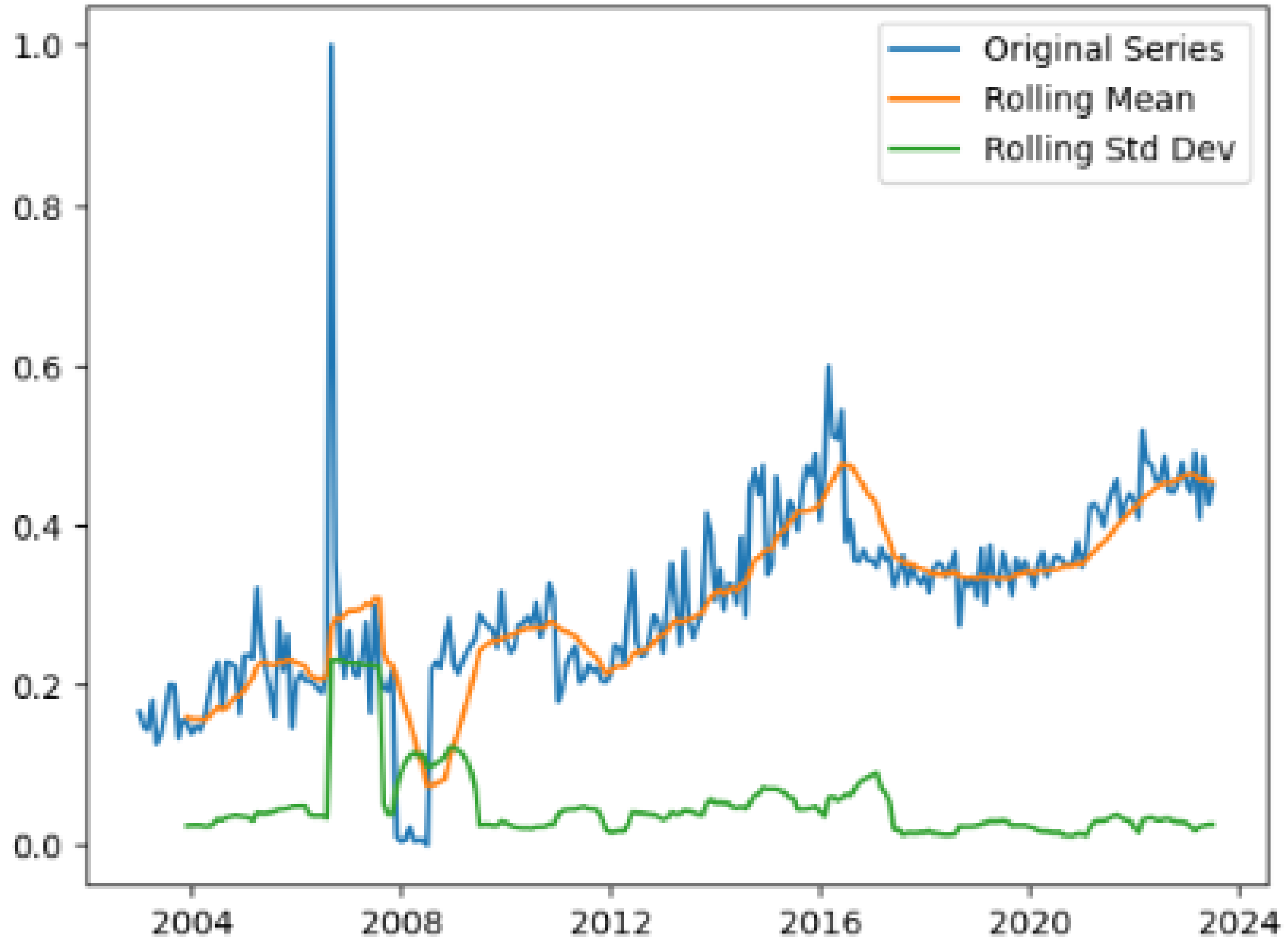
```
display(grouping_cnt)
```

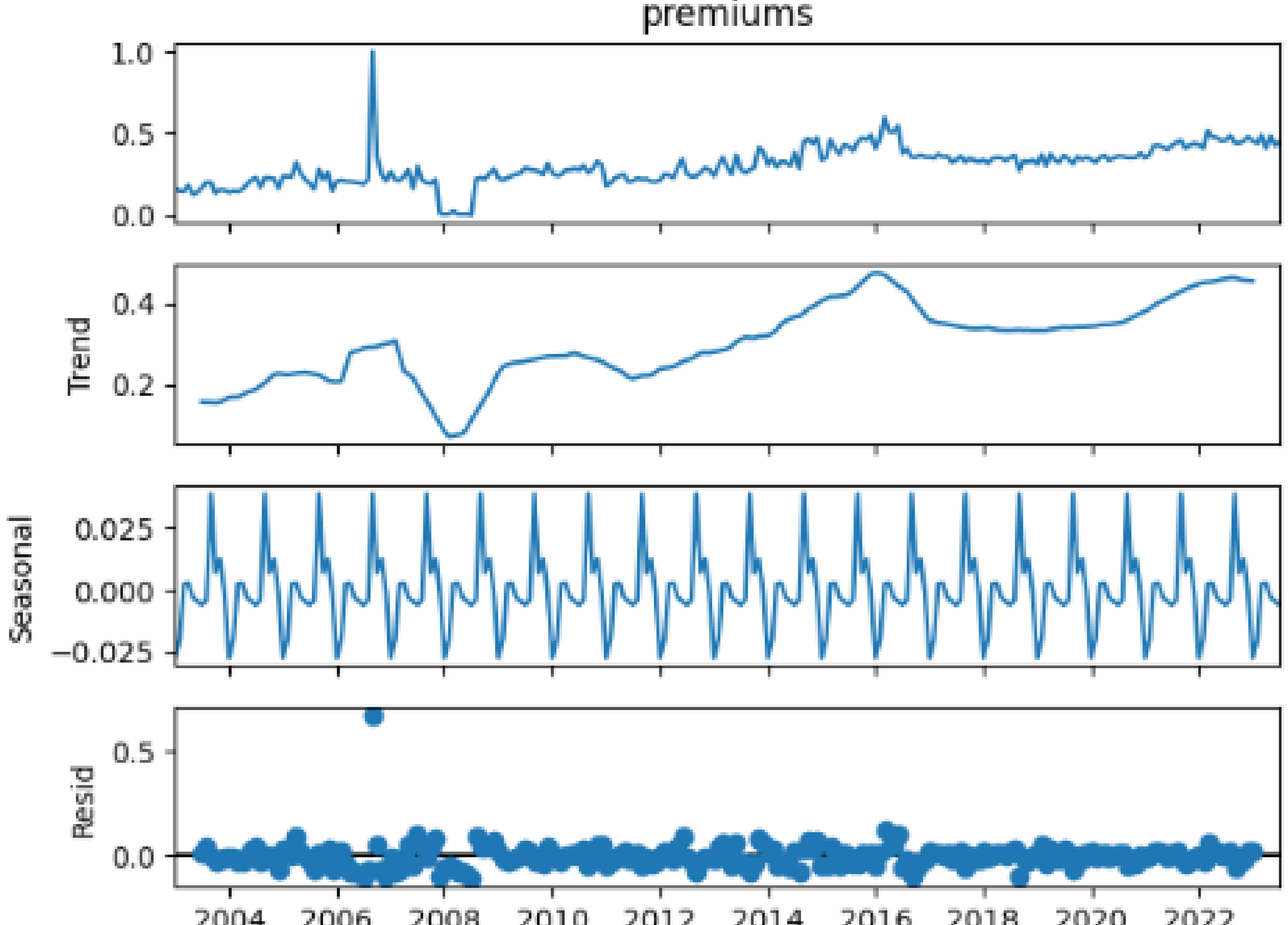
company_code	product	state	0
1007	0588 - DPVAT	AC	132
		AL	132
		AM	132
		AP	132
		BA	132
...	...	...	...
9938	1392 - VGBL/VAGP/VRGP/VRSA/VRI	RS	142
		SC	142
		SE	25
		SP	143
		TO	6

109448 rows × 1 columns


Boxplot of premiums
1e6
premiums
1.2
1.0
0.8
0.6
0.4
0.2
0.0


Histogram & KDE Plot
company_code
premiums
claims
claim_premium_ratio
Count
250
200
150
100
50


Original Series
Rolling Mean
Rolling Std Dev
1.0
0.8
0.6
0.4
0.2
0.0
2004
2008
2012
2016
2020
2024


premiums
1.0
0.5
0.0
Trend
0.4
0.2
Seasonal
0.025
0.000
−0.025
Resid
0.5
0.0
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022

Time Series Challenges & Solutions

📈 Observations:

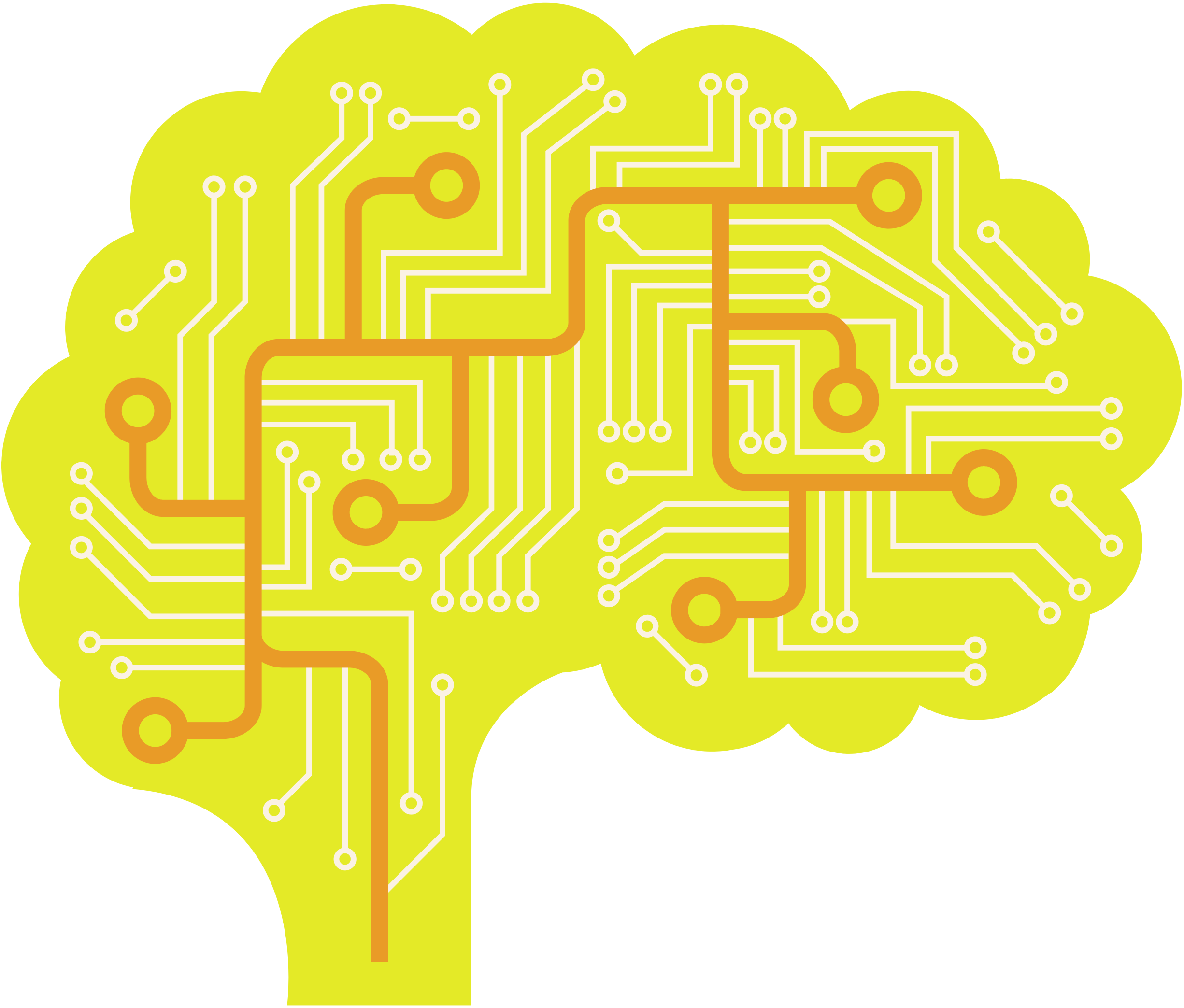
Data is non-stationary (changing over time).
Seasonal trends impact predictions.

🛠️ Techniques Applied:

- Differencing to remove trends
- Seasonal decomposition to identify hidden patterns
- ADF Test to confirm stationarity

Model Selection & Training

ML Models Implemented



01	**SARIMA**	Best for seasonal data
02	**Auto ARIMA**	Finds optimal parameters automatically
03	**Facebook Prophet**	Handles trends & seasonality efficiently

SARIMA Model Insights

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                    y   No. Observations:                247
Model:             SARIMAX(0, 1, 1)x(1, 0, [], 12)   Log Likelihood                410.721
Date:                             Mon, 03 Feb 2025   AIC                          -815.441
Time:                                     17:28:17   BIC                          -804.925
Sample:                                 01-01-2003   HQIC                         -811.207
                                      - 07-01-2023
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.4359      0.046     -9.564      0.000      -0.525      -0.347
ar.S.L12       0.1467      0.057      2.563      0.010       0.035       0.259
sigma2         0.0021      0.000     18.652      0.000       0.002       0.002
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):               141.66
Prob(Q):                              0.94   Prob(JB):                         0.00
Heteroskedasticity (H):               0.24   Skew:                             0.24
Prob(H) (two-sided):                  0.00   Kurtosis:                         6.69
```

ARIMA Model Insights

```
==============================================================================
Dep. Variable:               premiums   No. Observations:                  247
Model:                ARIMA(1, 1, 1)   Log Likelihood                 408.359
Date:               Mon, 03 Feb 2025   AIC                           -810.718
Time:                       17:28:17   BIC                           -800.202
Sample:                   01-01-2003   HQIC                          -806.484
                        - 07-01-2023
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1141      0.132      0.863      0.388      -0.145       0.373
ma.L1         -0.5424      0.119     -4.557      0.000      -0.776      -0.309
sigma2         0.0021      0.000     18.619      0.000       0.002       0.002
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):               141.33
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               0.26   Skew:                             0.14
Prob(H) (two-sided):                  0.00   Kurtosis:                         6.70
===================================================================================
```

Prophet Model Insights

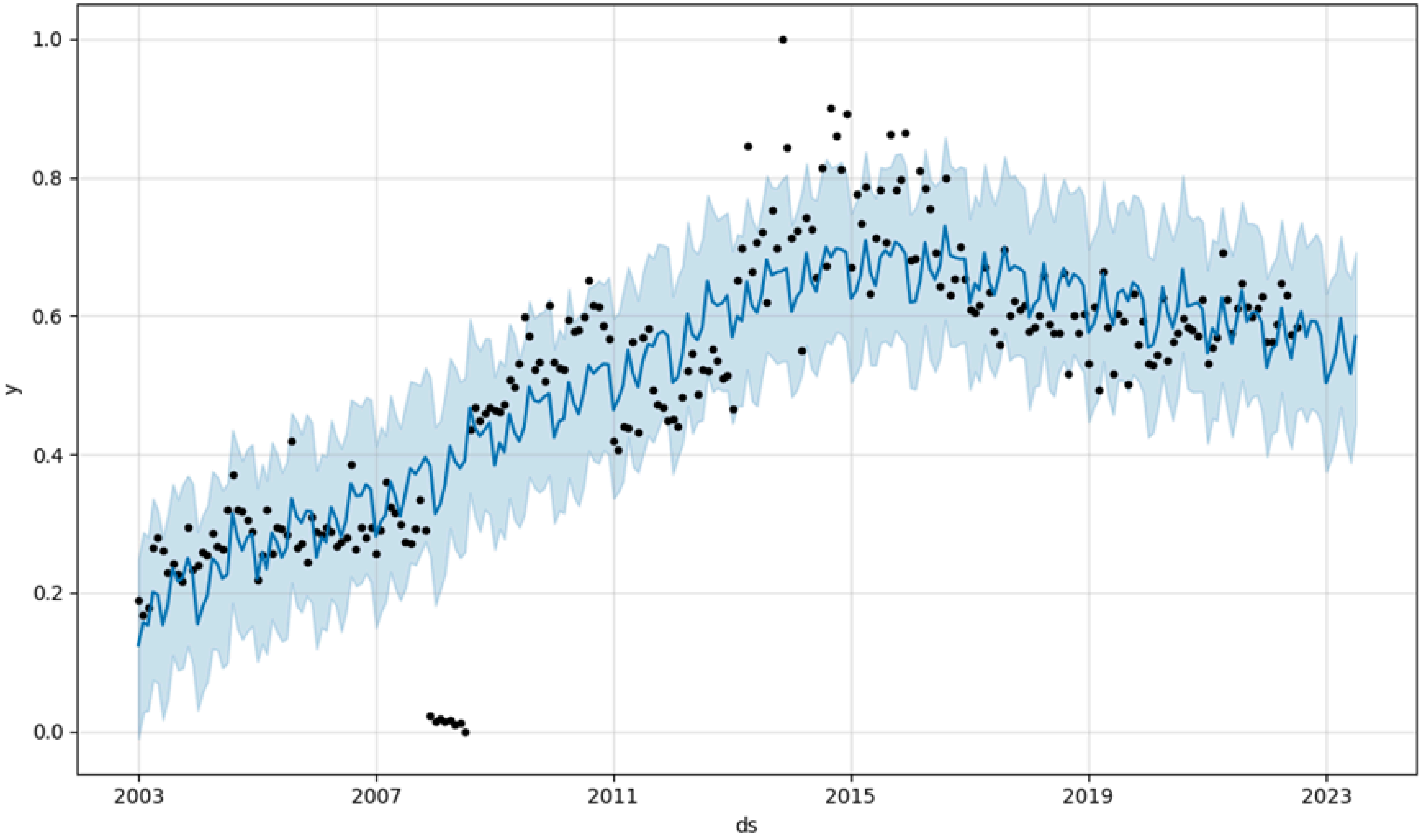


Fig.Checking outliers with Prophet

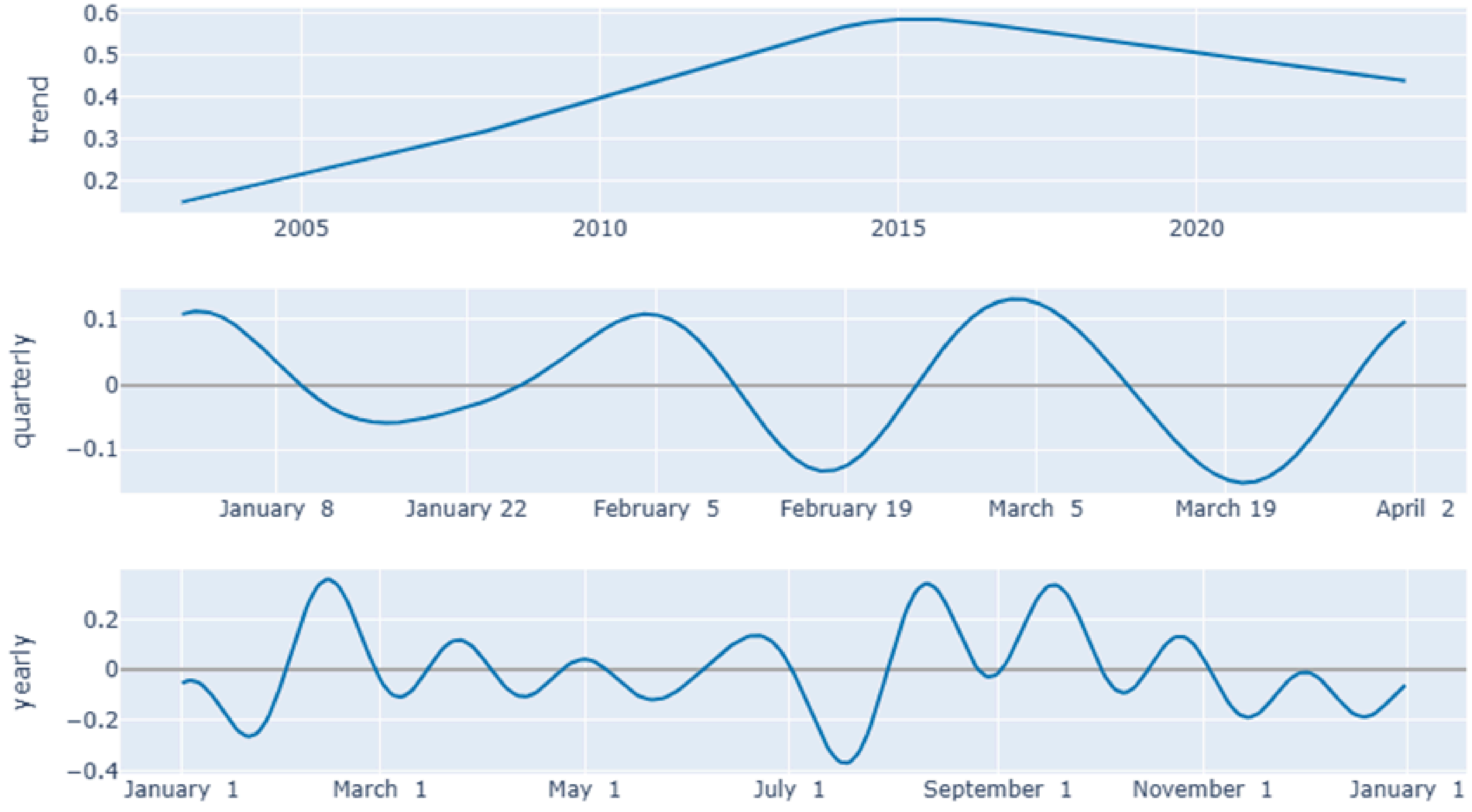


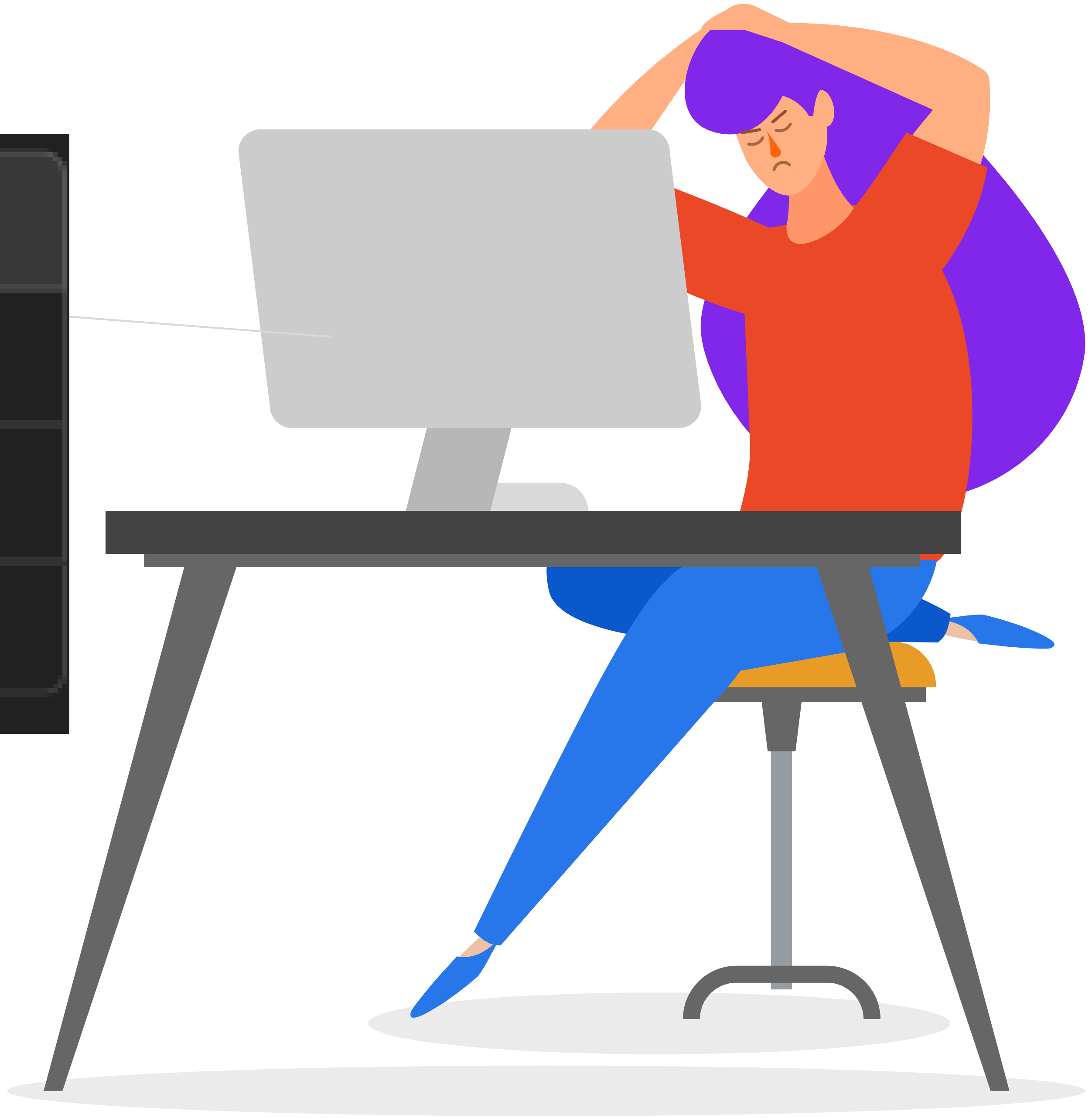
Fig. Plots of yearly, quarterly trends

Performance Metrics

Performance Metrics

Model	AIC	BIC	RMSE	MAPE
SARIMA	-815.4	-804.9	-	-
Auto ARIMA	-810.7	-800.2	-	-
Prophet	-	-	1476.45	4.74%

🏆 **Model Chosen: Facebook Prophet** ✅



Model Comparisons

SARIMA Model

Results

- AIC = -815.441, BIC = -804.925
- Captures seasonal trends well
- Low residual variance, indicating a strong fit

Limitations

- Requires manual tuning of parameters
- Sensitive to missing data

Auto ARIMA Model

Results

- AIC = -810.718, BIC = -800.202
- Residual analysis shows volatility in premiums

Limitation

- Struggles with sudden market changes

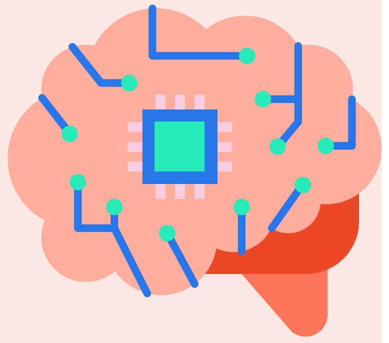
Facebook Prophet Model

Why Prophet?

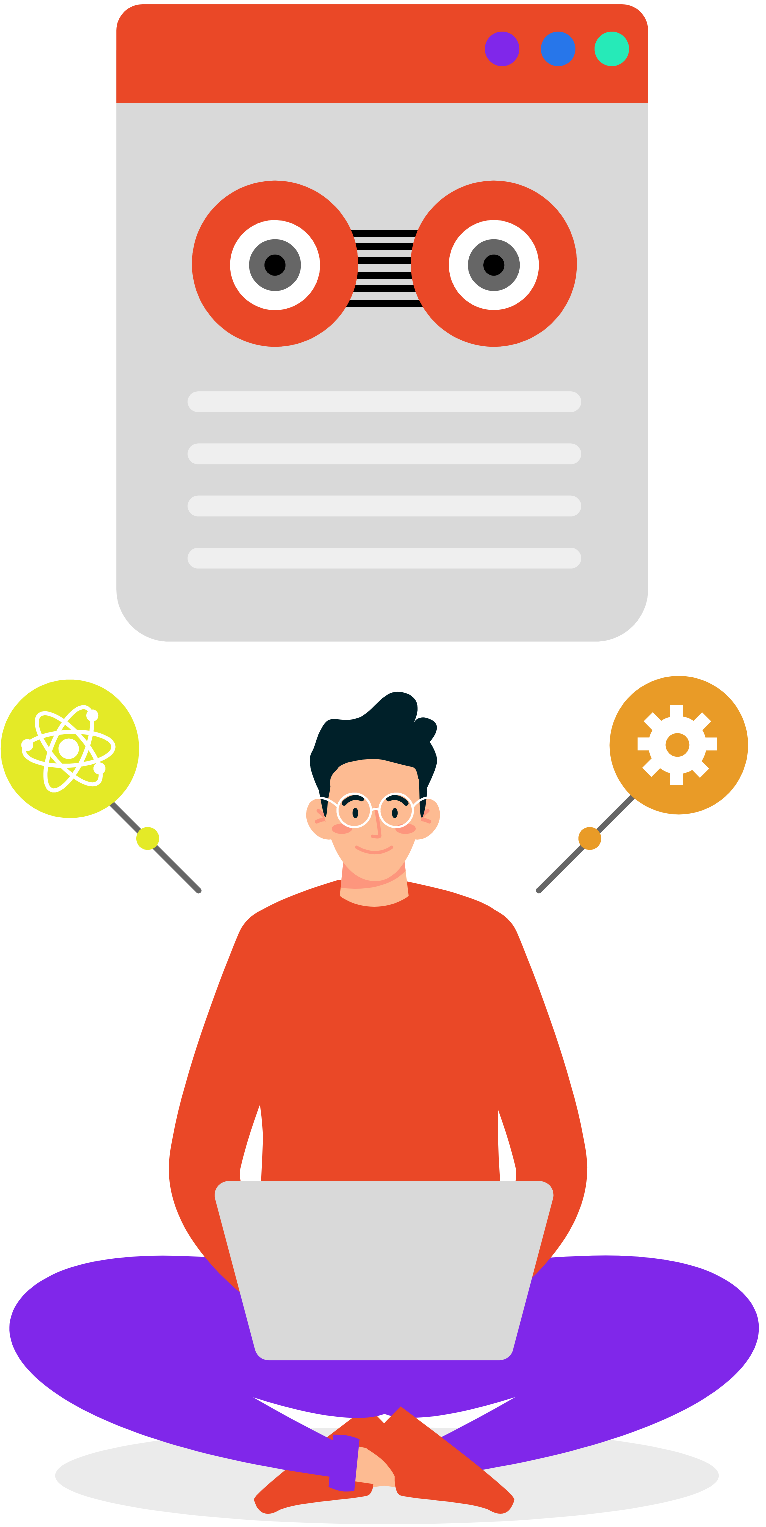
- Handles missing data well
- Captures yearly & quarterly seasonality
- Provides confidence intervals

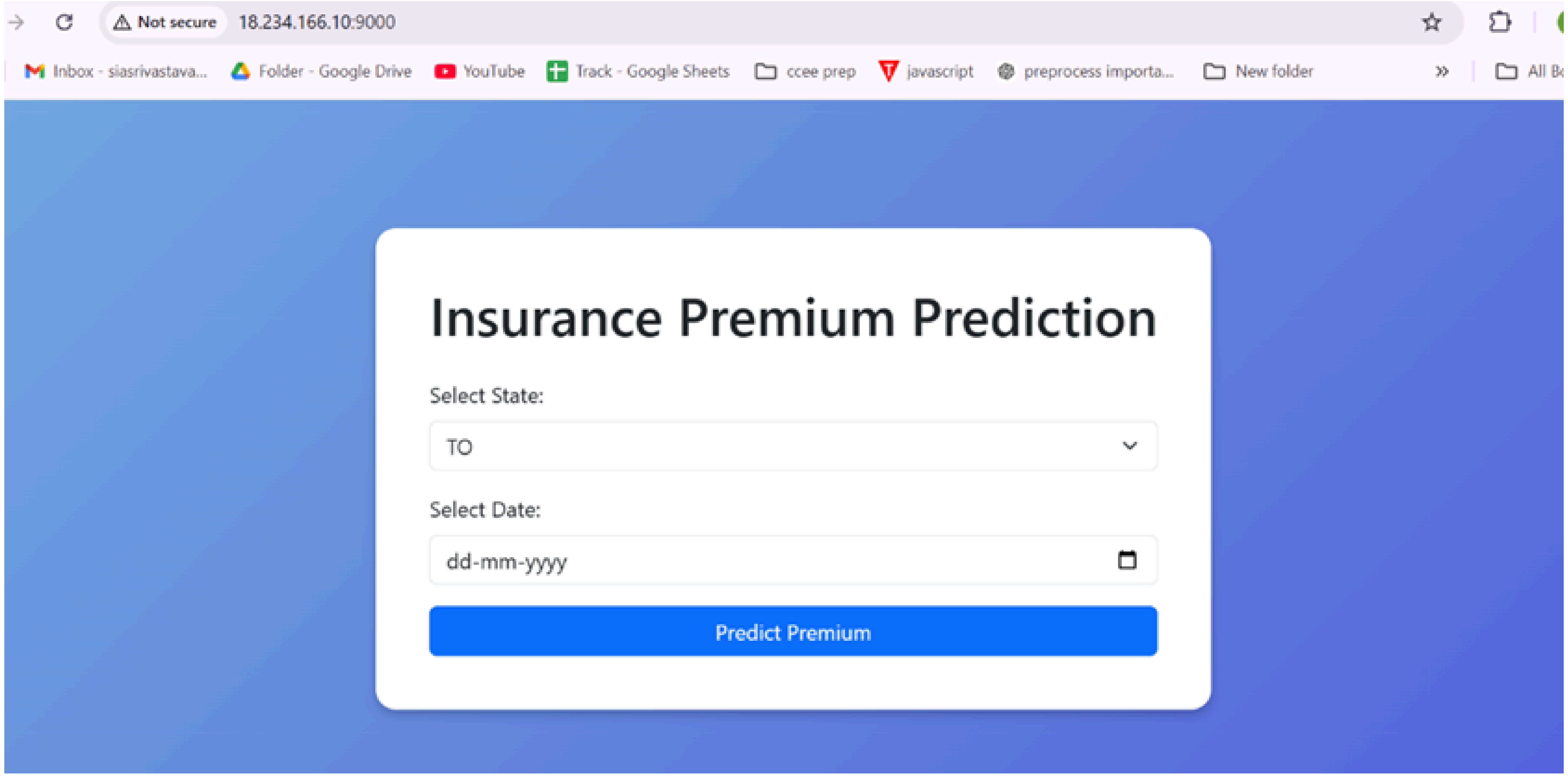
Results

- MAE: 28.10
- RMSE: 1476.45
- MAPE: 4.74%

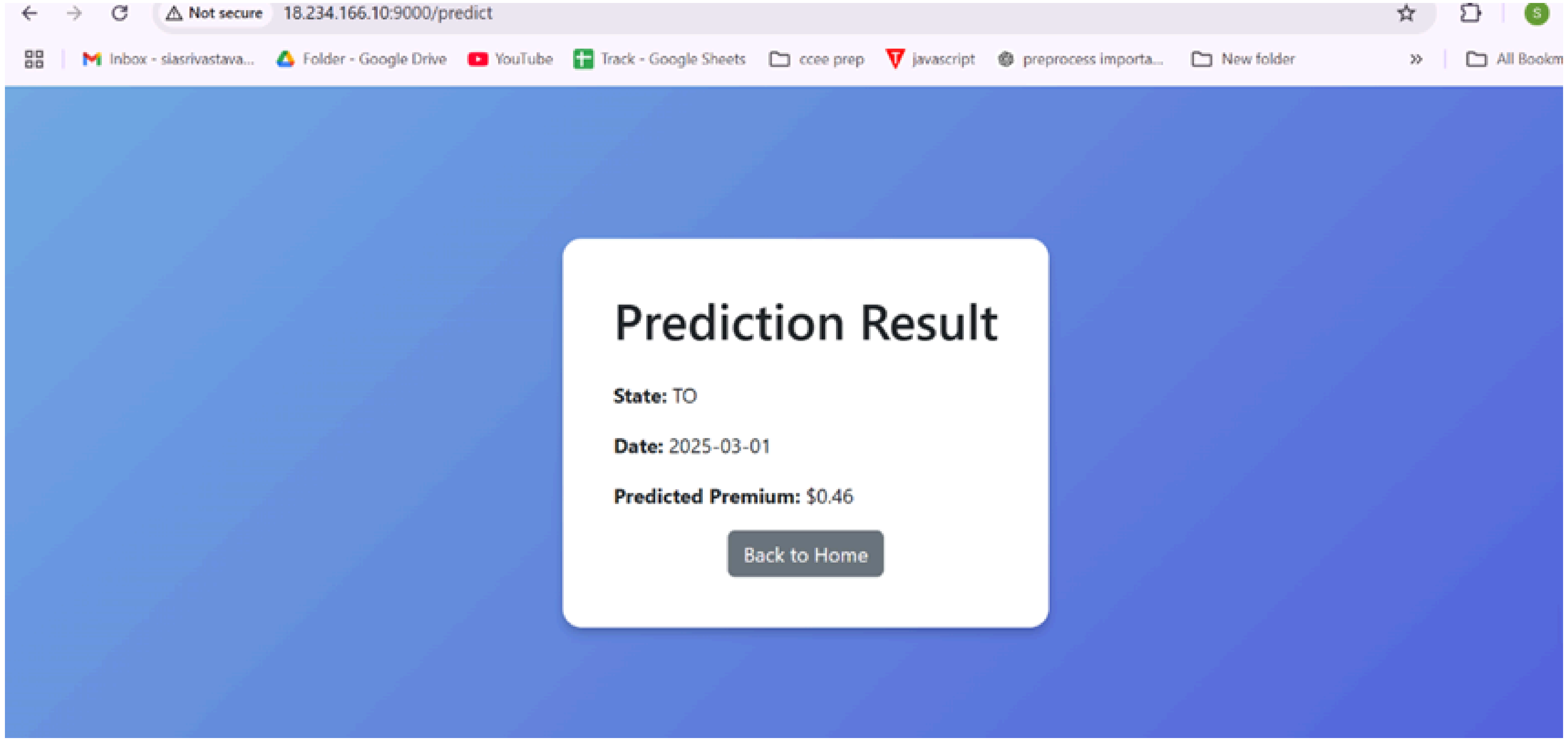


User Interface & Deployment

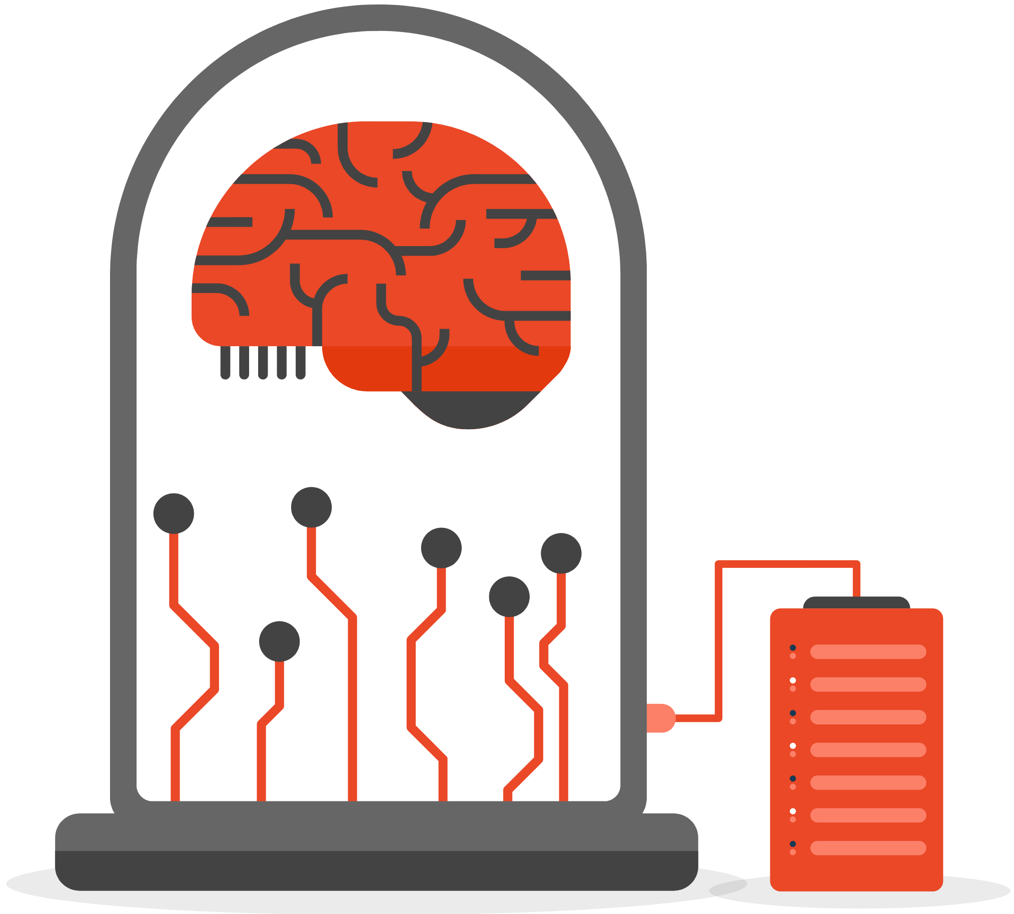




Main Landing page


Not secure
18.234.166.10:9000/predict
Inbox - siasrivastava...
Folder - Google Drive
YouTube
Track - Google Sheets
ccee prep
javascript
preprocess importa...
New folder
All Bookm
Prediction Result
State: TO
Date: 2025-03-01
Predicted Premium: $0.46
Back to Home

Result page



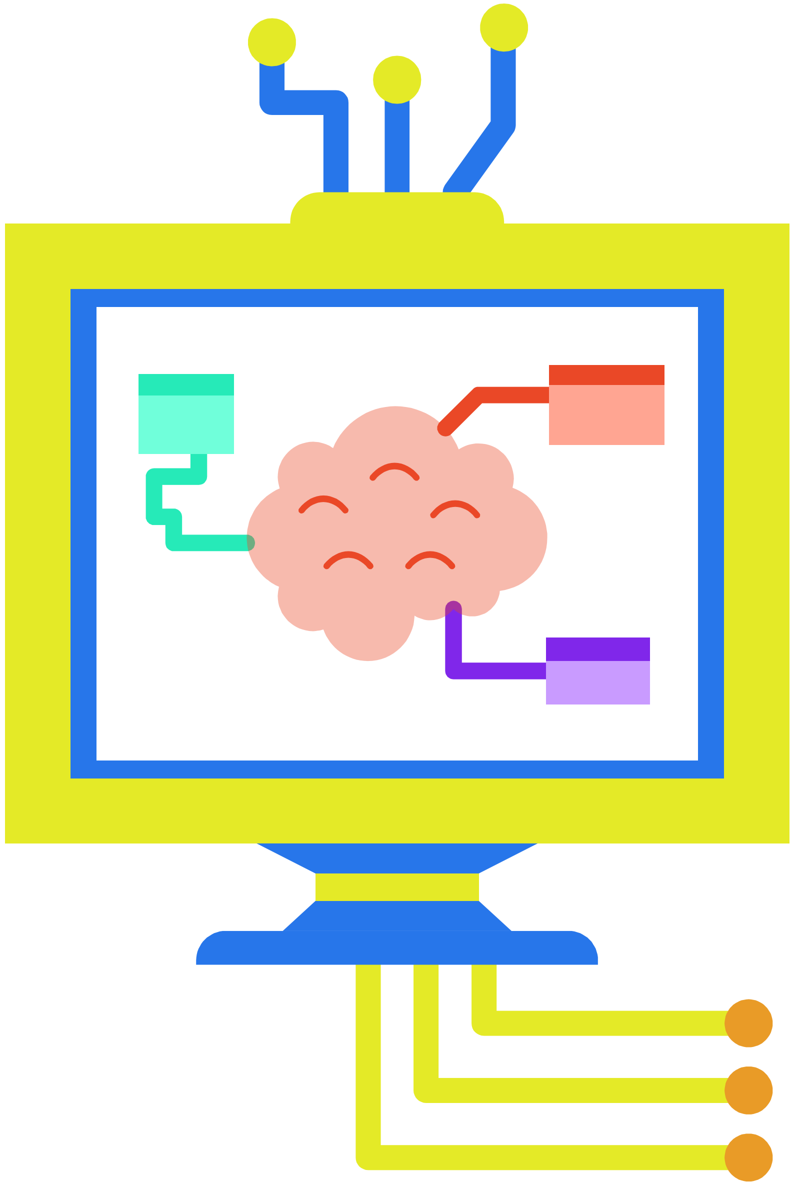
Requirements & Specifications

Hardware Requirement

- 500 GB hard drive (Minimum)
- 8 GB RAM (Minimum)
- PC x64-bit CPU

Software Requirement

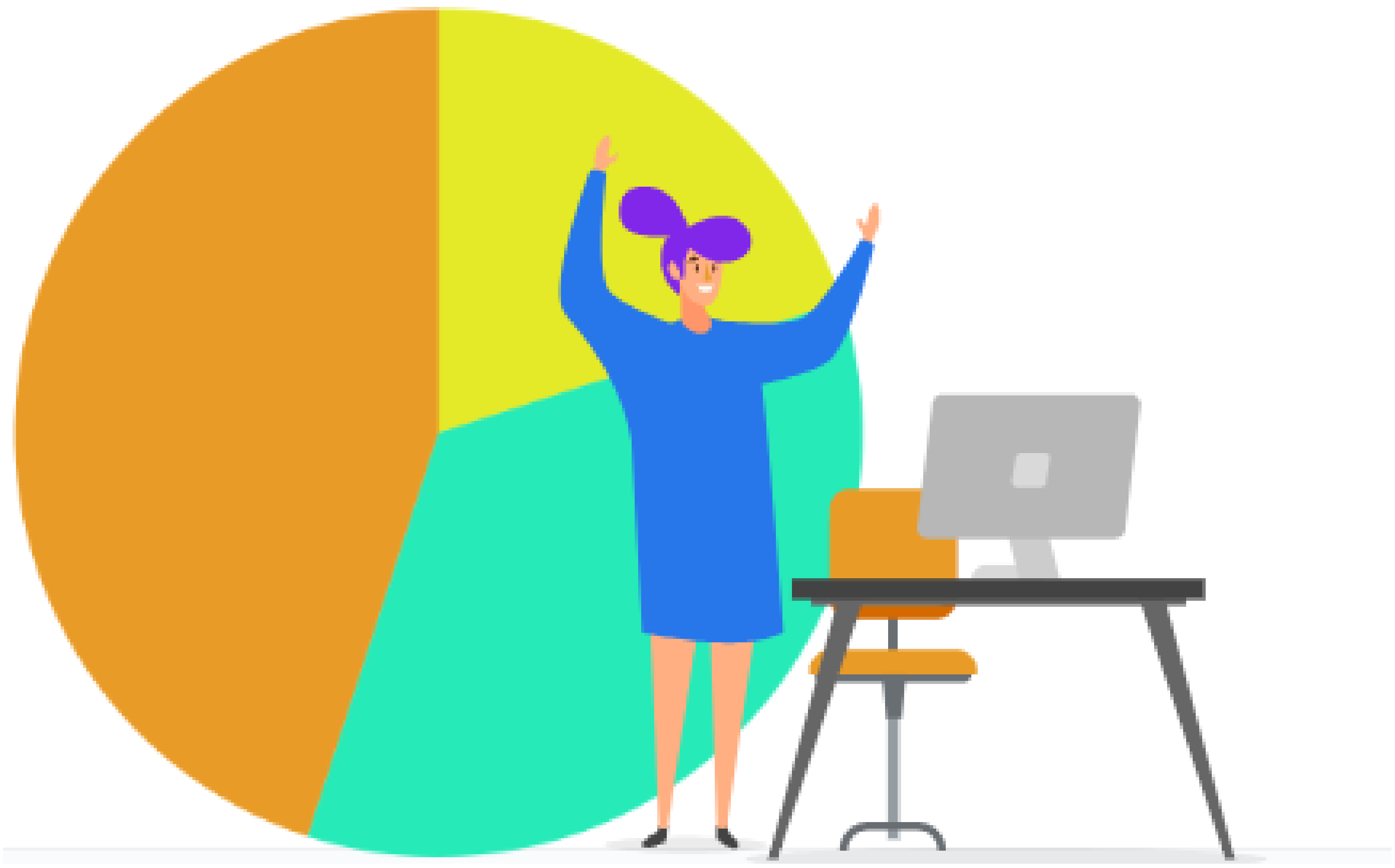
- Windows/Mac/Linux
- Python-3.9.10
- VS Code/Anaconda/Google Colab/Jupyter
- Python Extension for VS Code



Libraries

· Flask=1.1.1
· prophet=1.1.5
· numpy=1.9.2
· scipy=0.15.1
· scikit-learn=0.18
· matplotlib=1.4.3
· pandas=0.19
· Any Modern Web Browser
AWS Cloud = EC2 service

Outcomes

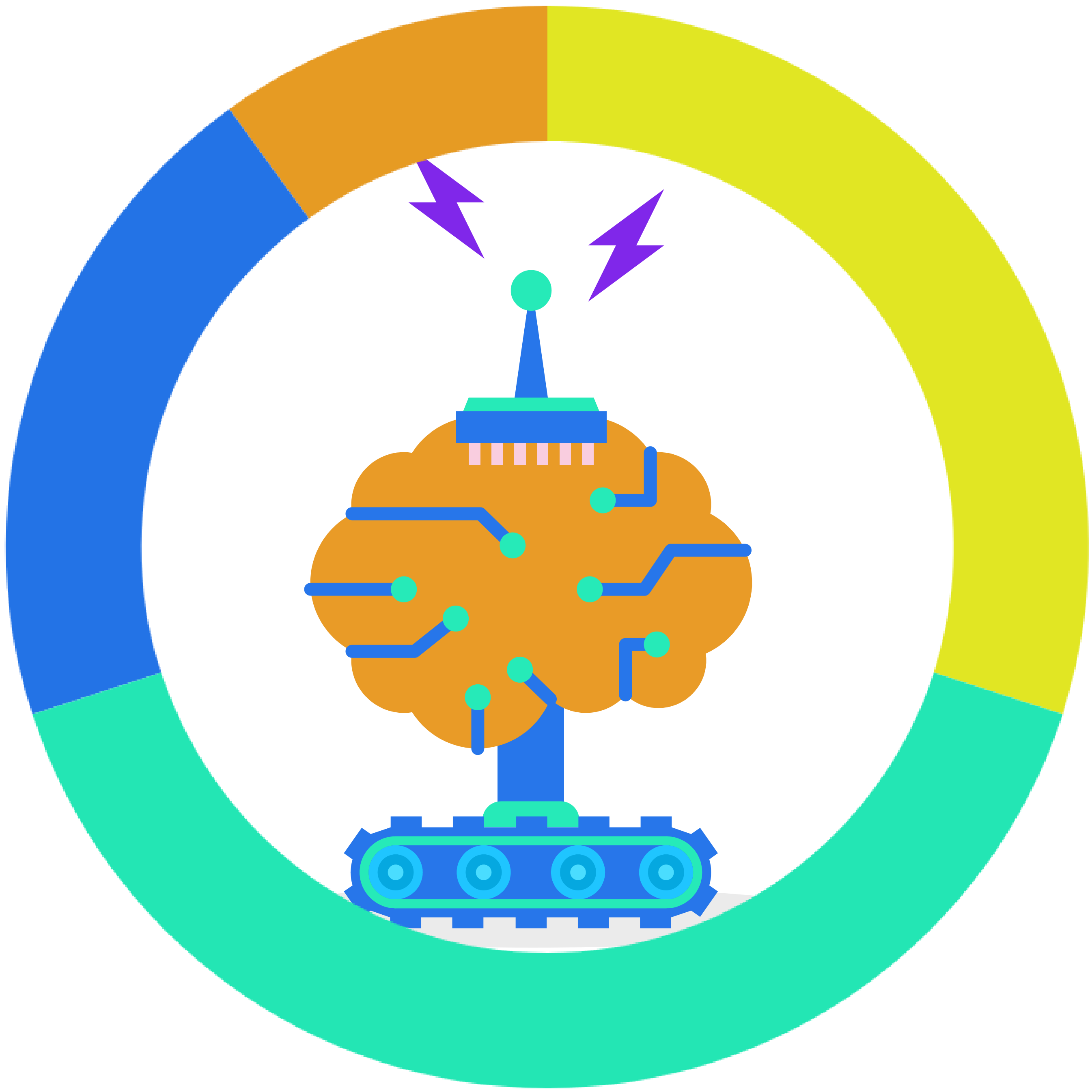


✅ Successfully implemented insurance premium prediction.

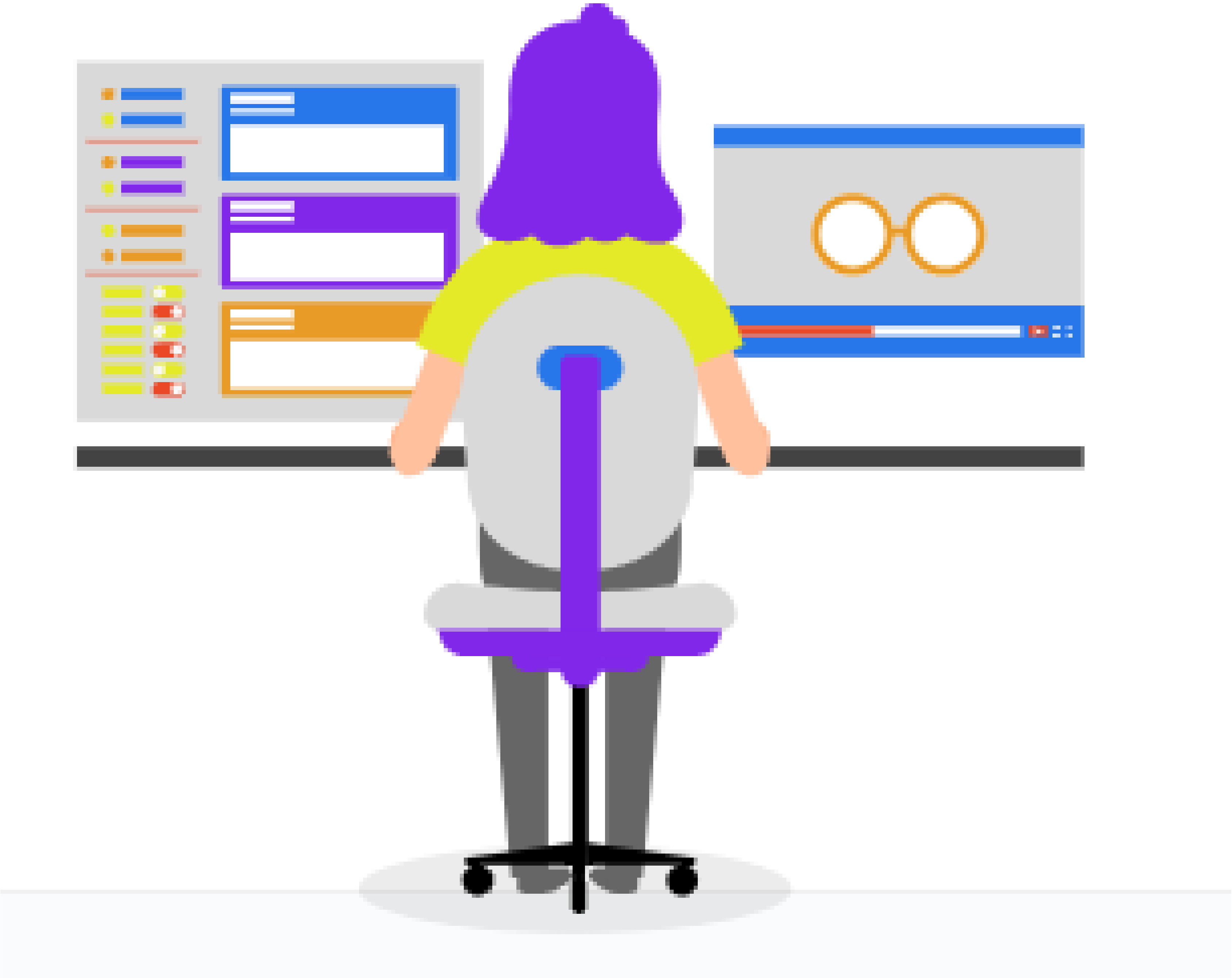
✅ Facebook Prophet performed best for time-series forecasting.

✅ Web app enables real-time premium estimation.

Conclusion



- This project implemented a predictive modeling system using Facebook Prophet for time series forecasting, identifying key patterns to enhance decision-making.
- Integration with Flask provided an interactive interface for easy model access, while deployment on AWS EC2 ensured scalable, real-time predictions with minimal latency.



Thank You