

# APP:

## Exploratory:

**Understand & Predict Salaries**

**Data Exploration**

**Salary Prediction**

**Customize Visualization**

Choose Cutoff Value for Country Count  
100 1,000 10,000

Choose Cutoff Value for Salary Range  
1,000 10,000 100,000

Select Countries to Include  
Choose an option

## Exploratory Data Analysis

### Introduction

### A First Look at Our Data

We have access to an annual survey by Stack Overflow where we can find more than 67K software developers' answers. Part of the survey includes salary, experience, country, full/part-time employee, language, demographic information and so on. After processing and cleaning the data, here we have 50 samples of what our final dataset looks like. In the following sections you will be able to deep dive into the specifics of this dataset, and you may be able to find insightful patterns.

Country	EdLevel	YearsCode/ro	Salary
59,247 Other	Bachelor's degree	2	6,496
52,034 Other	Master's degree	7	41,256
9,913 United States of America	Less than a Bachelor's	3	65,000
61,561 United States of America	Master's degree	5	200,000
60,593 Slovenia	Master's degree	4	48,624
70,781 Iran, Islamic Republic of...	Bachelor's degree	10	85,104
11,270 United States of America	Bachelor's degree	14	115,000
50,671 United Kingdom of Great Britain and Northern Ireland	Master's degree	25	64,076
18,333 Spain	Bachelor's degree	2	41,254
6,542 Germany	Master's degree	10	110,680

### Distributions

The main insights to take from the collection of plots within this section are related to how all the different variables are distributed. That is, the counts of records of each particular variable. That can be also seen as the number of developers for each distinct value of a variable.

**Understand & Predict Salaries**

**Data Exploration**

**Salary Prediction**

**Customize Visualization**

Choose Cutoff Value for Country Count  
100 1,000 10,000

Choose Cutoff Value for Salary Range  
1,000 10,000 100,000

Select Countries to Include  
Choose an option

### Distributions

The main insights to take from the collection of plots within this section are related to how all the different variables are distributed. That is, the counts of records of each particular variable. That can be also seen as the number of developers for each distinct value of a variable.

#### Developers Distribution per Country

This first chart illustrates the distribution of developer counts by country. We can see that we have mostly data of software developers in USA. The next most common country after that is shown to be 'Others' which are all the aggregated countries with lower count than the set cut-off value. (It can be changed on the side bar)

Number of Developers

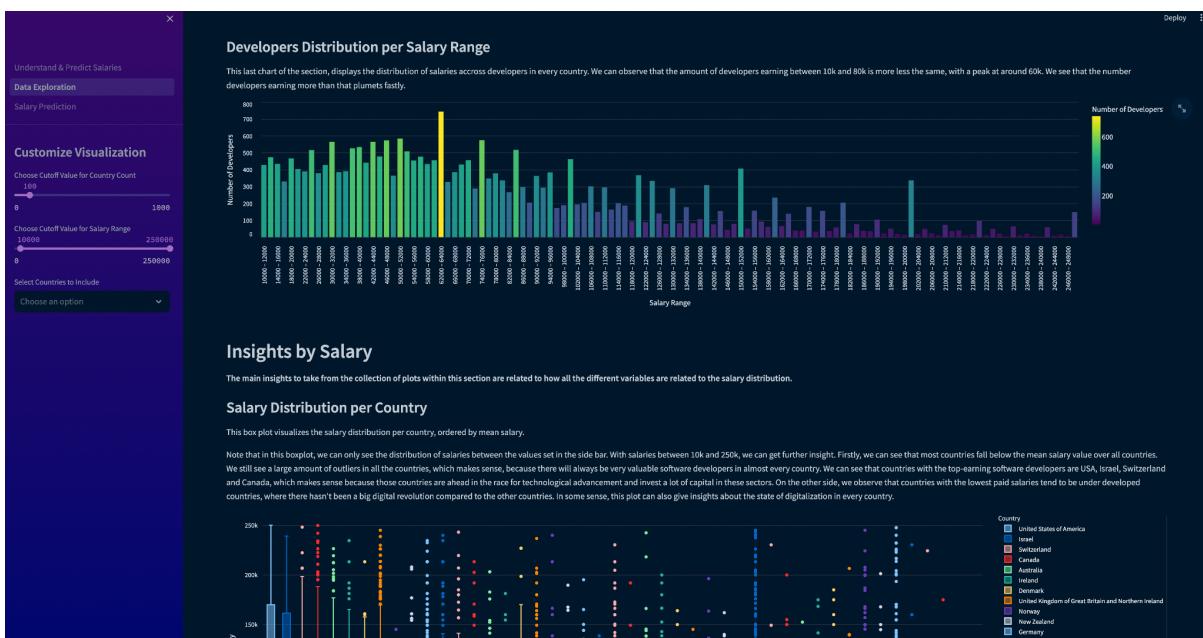
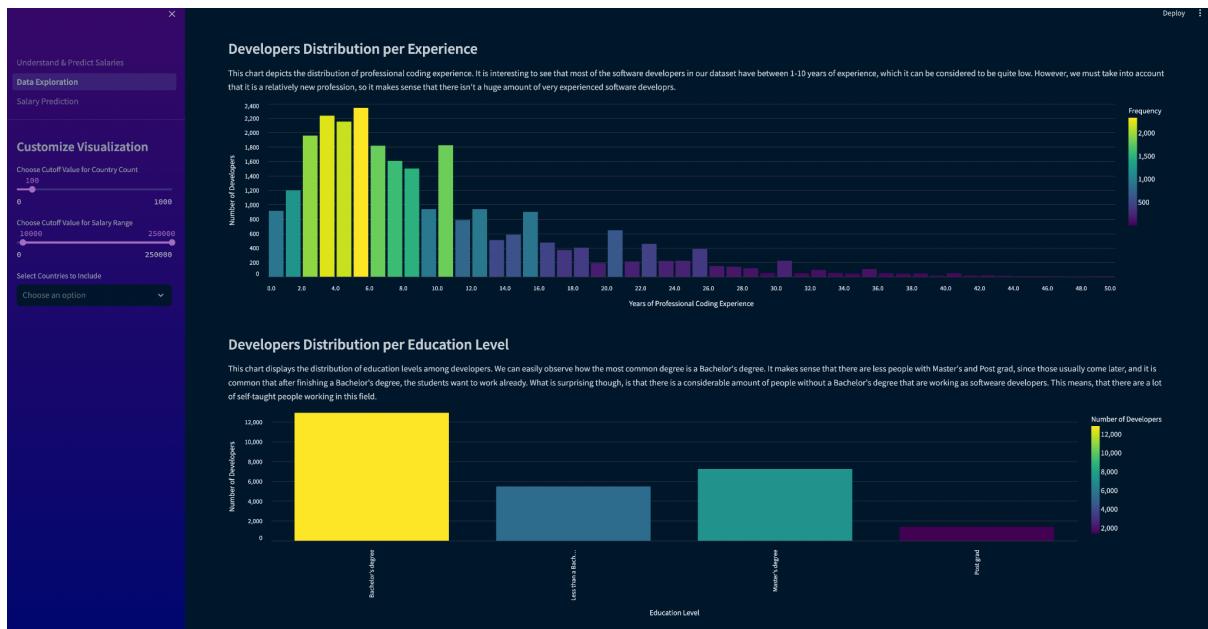
Country

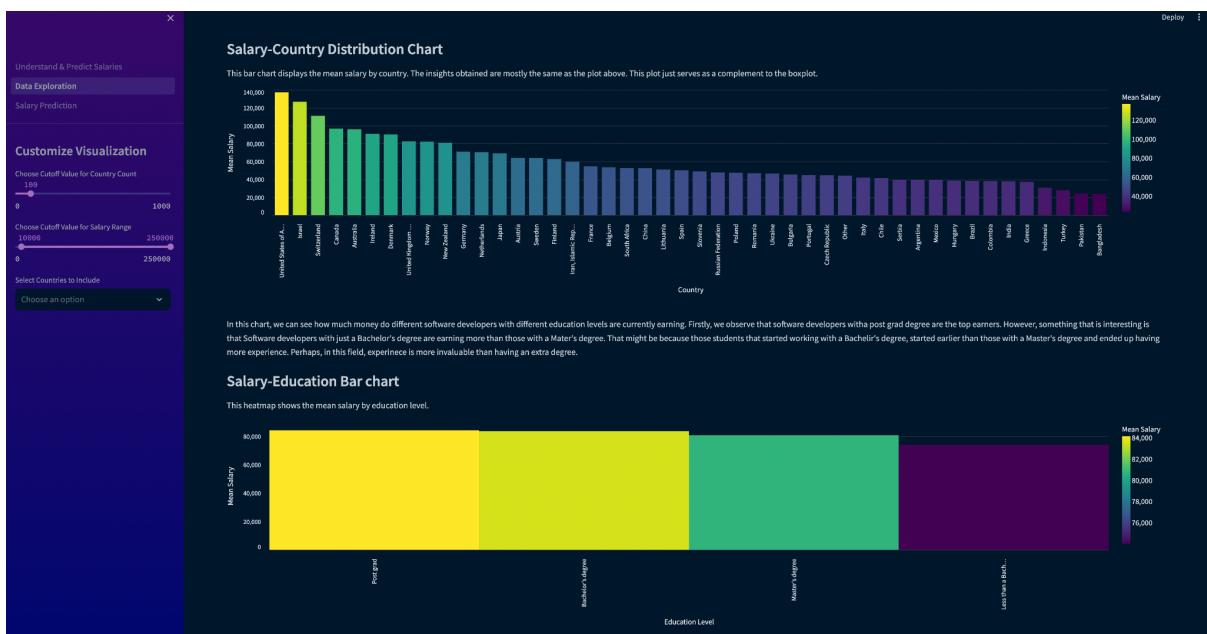
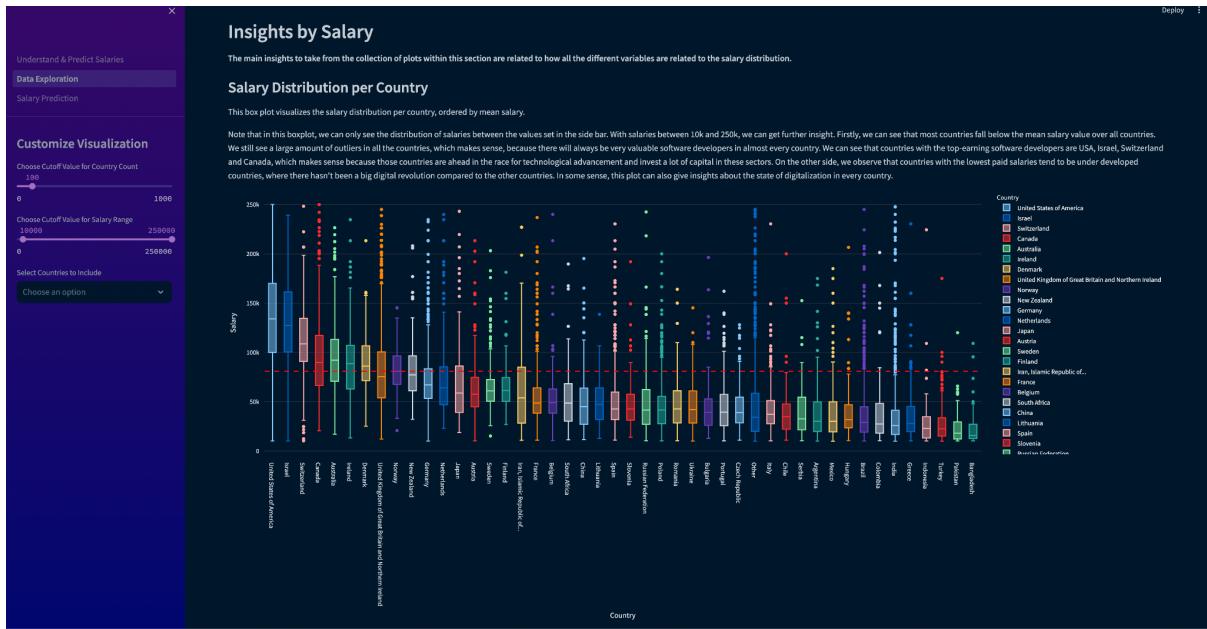
#### Developers Distribution per Experience

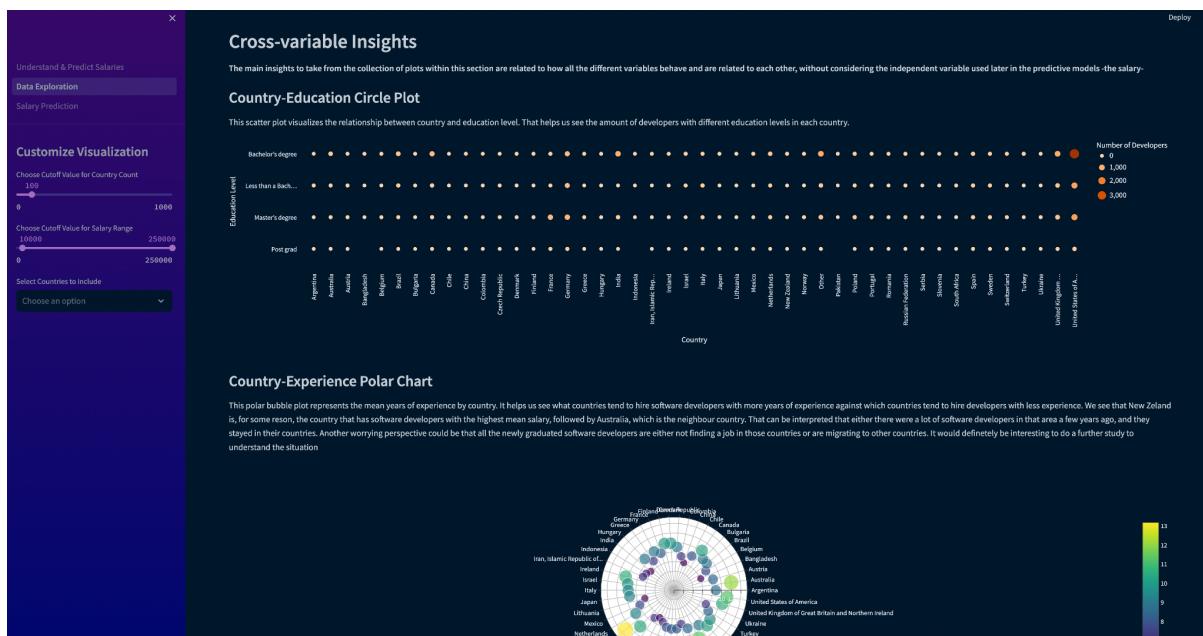
This chart depicts the distribution of professional coding experience. It is interesting to see that most of the software developers in our dataset have between 1-10 years of experience, which it can be considered to be quite low. However, we must take into account that it is a relatively new profession, so it makes sense that there isn't a huge amount of very experienced software developers.

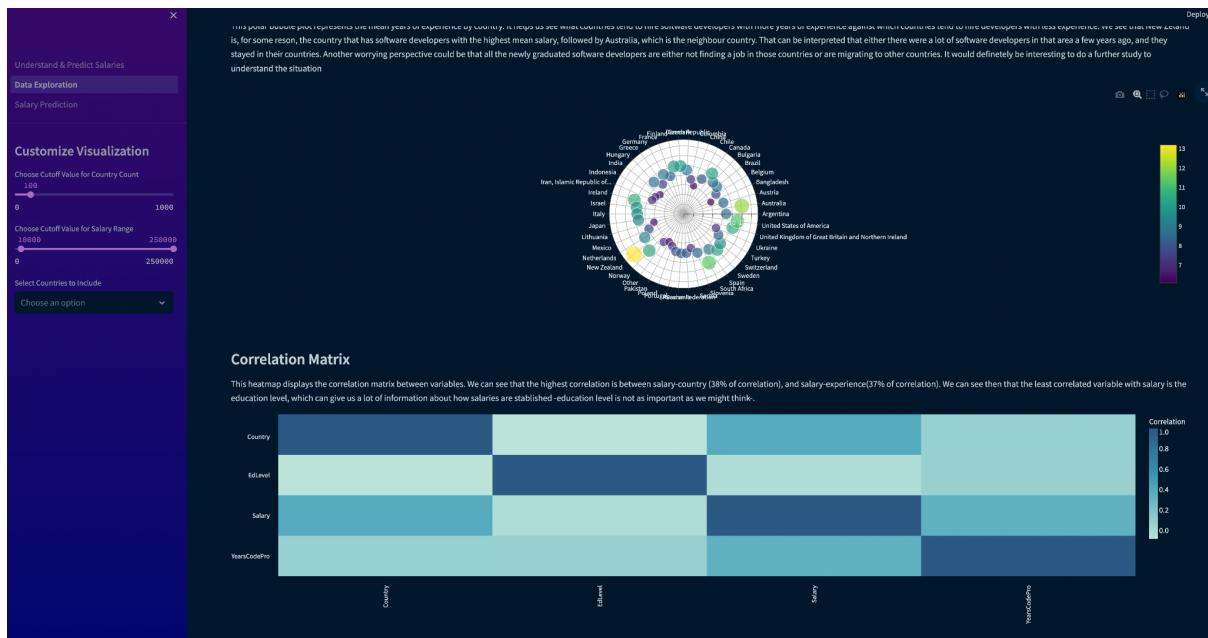
Number of Developers

Frequency

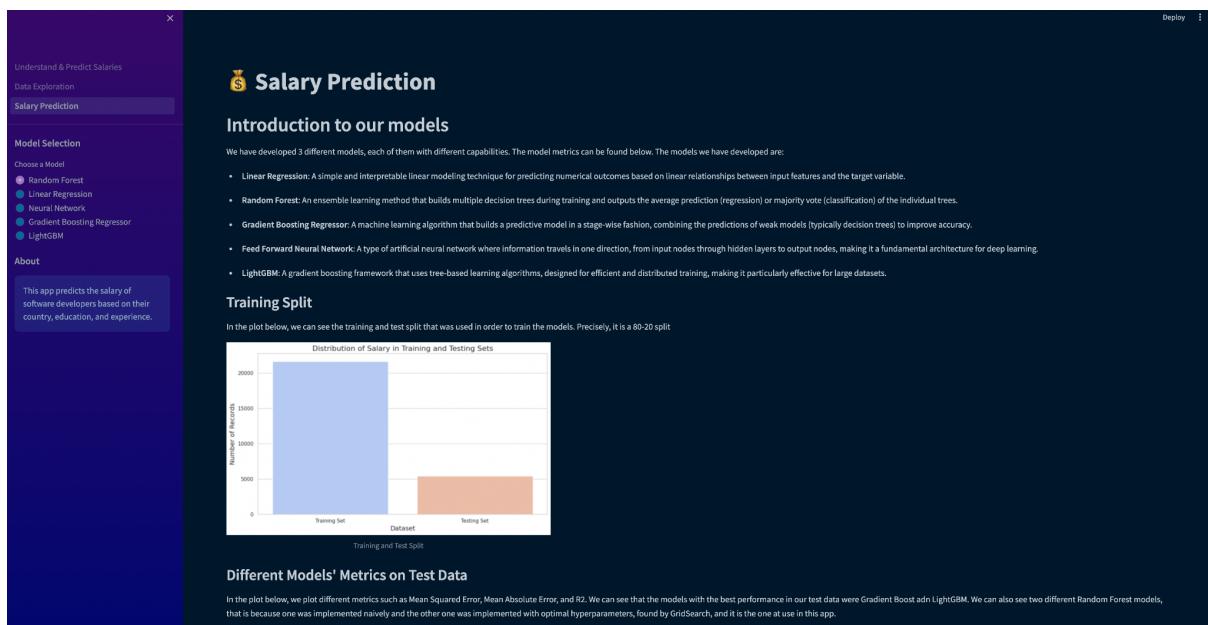


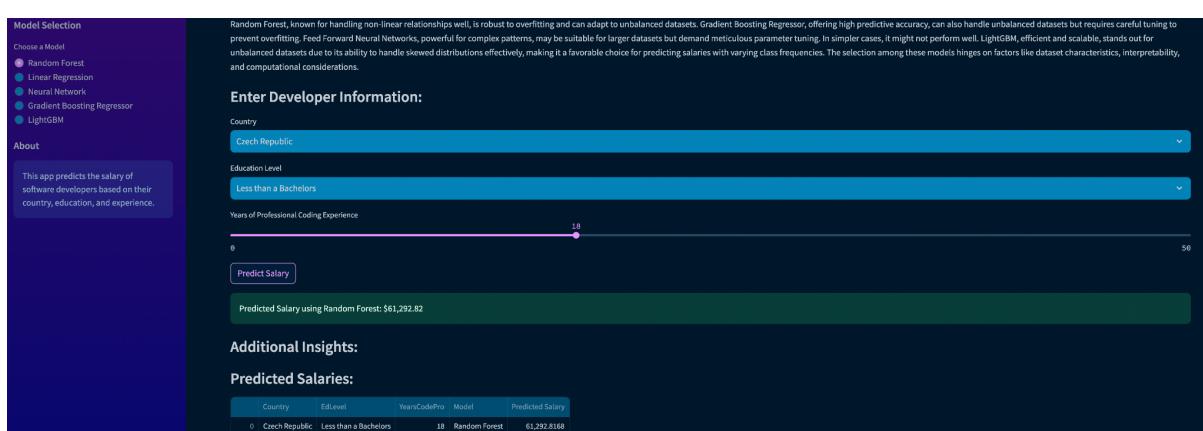
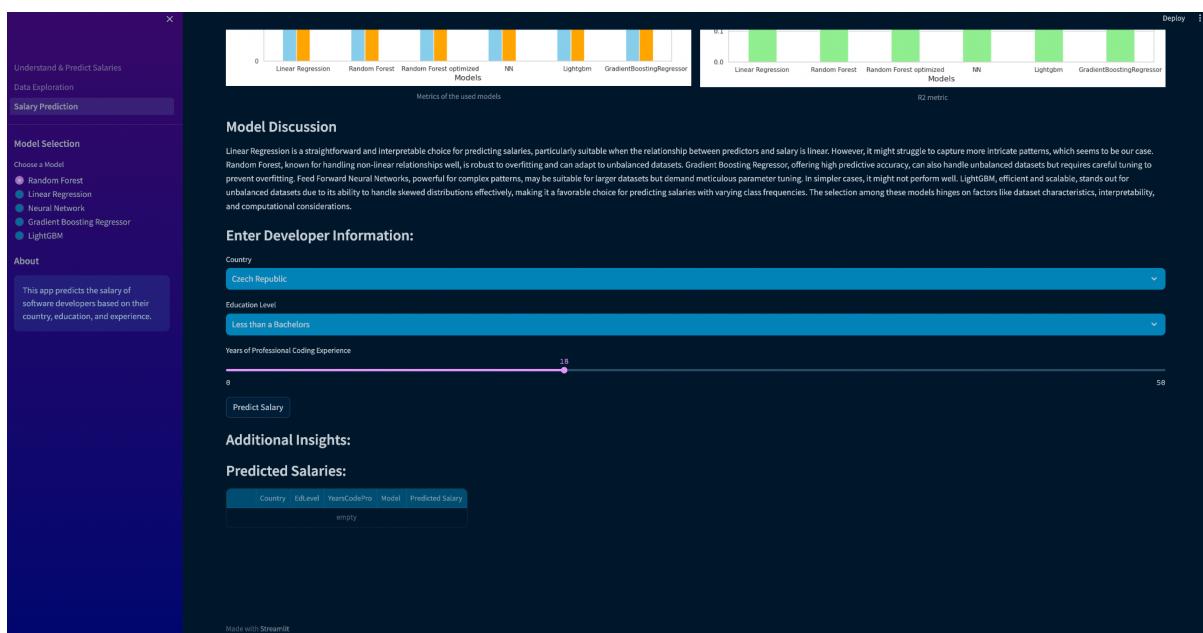
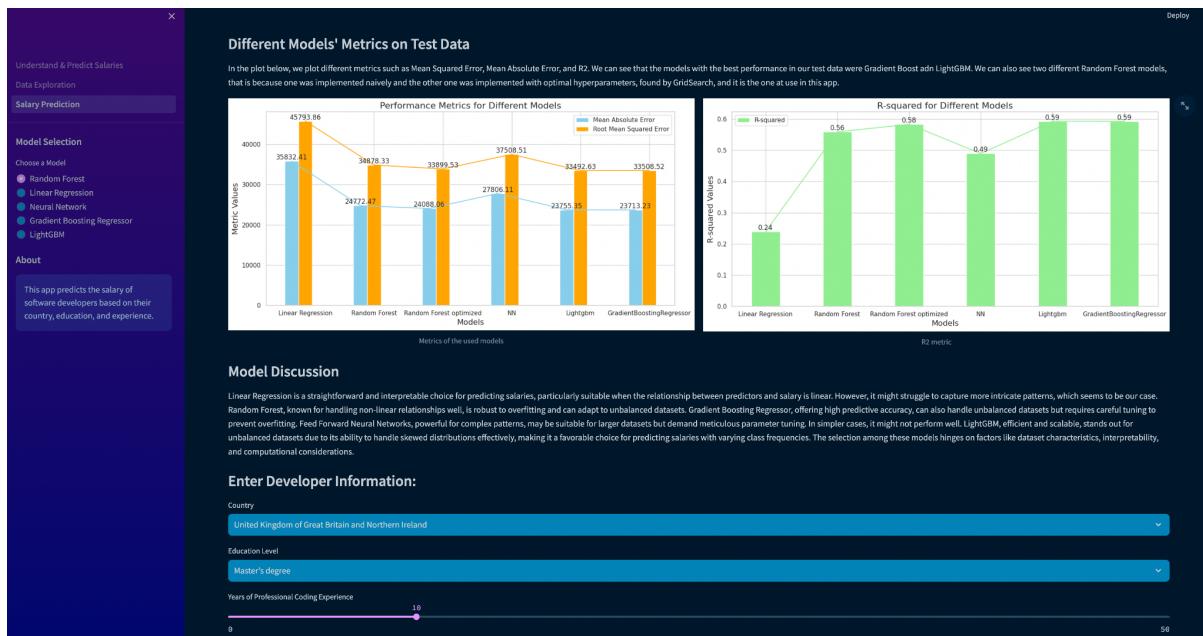






## Models:





**Model Selection**

Choose a Model

- Random Forest
- Linear Regression
- Neural Network
- Gradient Boosting Regressor
- LightGBM

About

This app predicts the salary of software developers based on their country, education, and experience.

unbalanced datasets due to its ability to handle skewed distributions effectively, making it a favorable choice for predicting salaries with varying class frequencies. The selection among these models hinges on factors like dataset characteristics, interpretability, and computational considerations.

**Enter Developer Information:**

Country

Israel

Education Level

Post grad

Years of Professional Coding Experience

0 12 56

**Predict Salary**

Predicted Salary using Gradient Boosting Regressor: \$166,063.57

**Additional Insights:**

**Predicted Salaries:**

	Country	EdLevel	YearsCodePro	Model	Predicted Salary
0	Czech Republic	Less than a Bachelors	18	Random Forest	61,292.0168
1	Israeli	Post grad	12	Gradient Boosting Regressor	166,063.5685