

# Salary Prediction

Regression Analysis

# Project Description

Main objective

What are your salary expectations?

Data Components

- \* Company Size
- \* Company
- \* Type of Ownership
- \* Job Title
- \* Sector
- \* Skill sets

Tools Used:

- Pandas
- SQL Spark
- Tableau
- sklearn

# Overview

Data Collection

Data Exploration

Regression Model

Conclusion

# Data Collection

Source:

- [Kaggle](#)
- Scraped from Glassdoor.com
- 2017 dataset



# Data Exploration

- Extract, Transform and Load- ETL
- Data Visualizaton- Tableau

# ETL



Finding  
Duplicates

Focusing on  
Columns Needed

Relevant columns for the  
analysis

Splitting and  
Deleting special  
characters

Cleaning up each column  
with special values or  
characters

Exploring each  
columns

Unique values of each  
column

# Questions

- What's the median pay for roles in the field of data?
- What's the top-paying job title in the data-driven domain?
- Which technical skills are the hottest in the job market?
- Which state has the highest paying data related jobs?
- Is my paycheck influenced by the company's size and industry sector?
- Does higher salary mean higher job satisfaction?



# 97k

Median Salary

# Python

Hottest technical skill

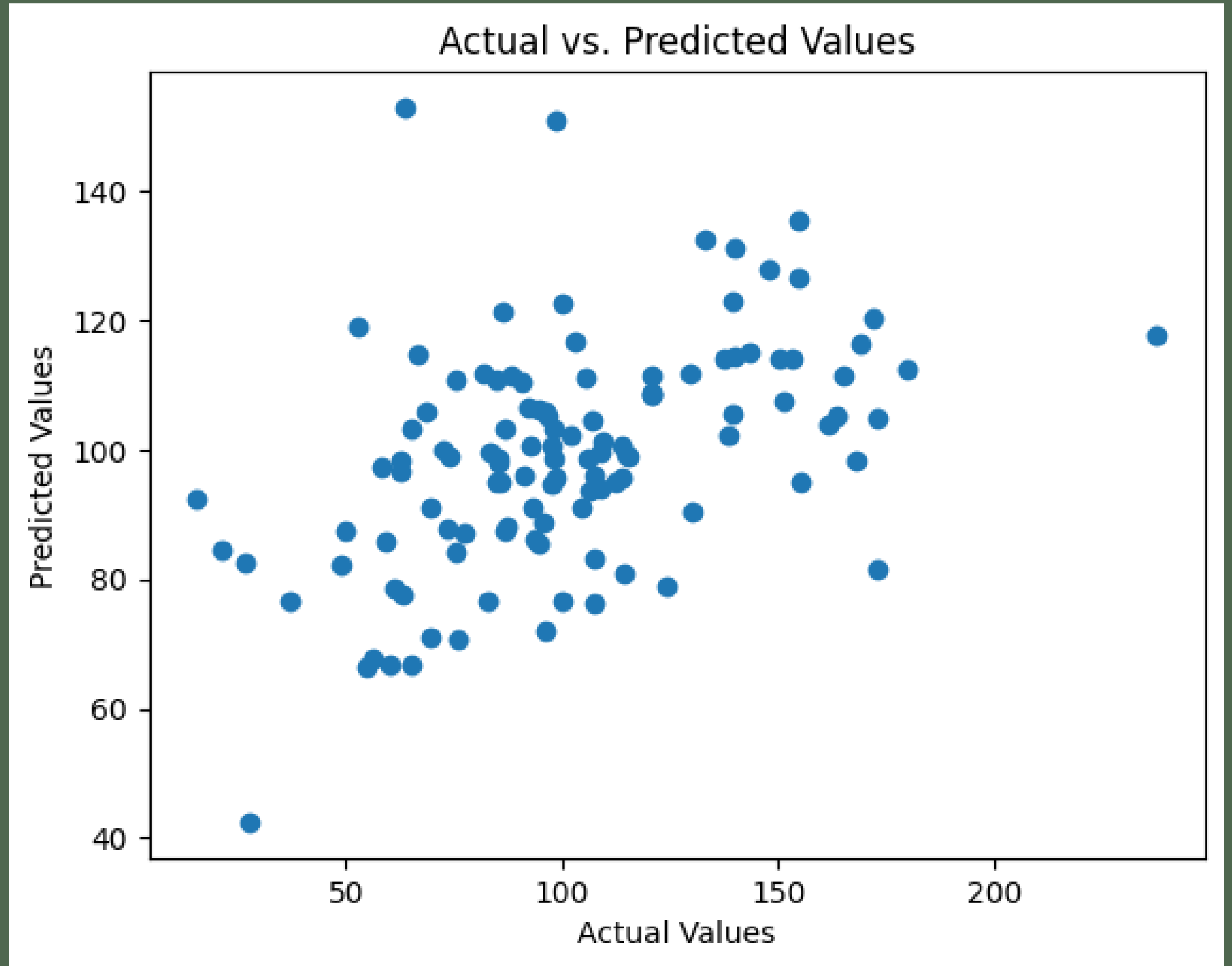
# California

Highest paying data related jobs



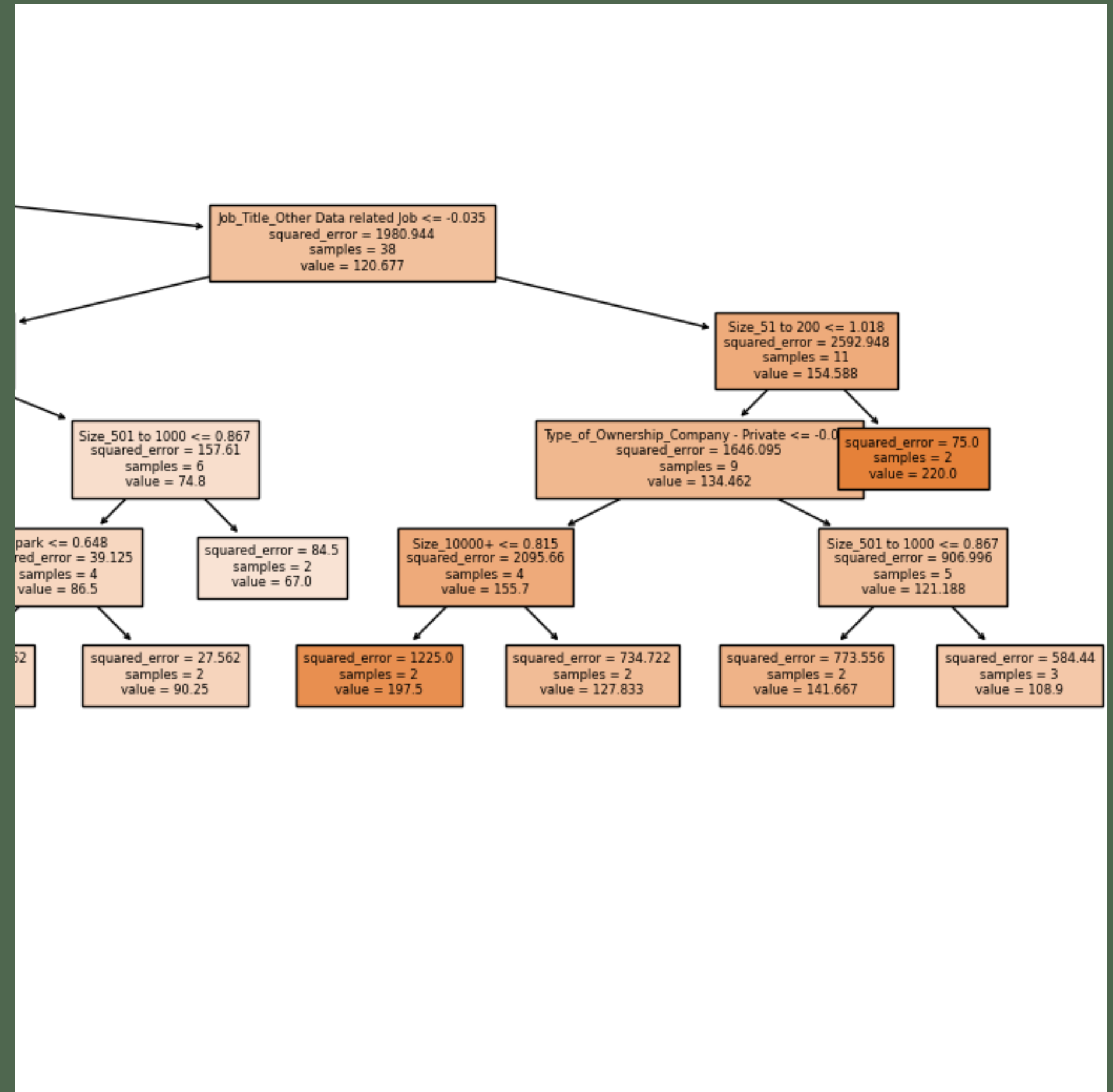
# Random Forest Regression Model

- Mean Squared Error: 1132.4634701623736
- Root Mean Squared Error: 33.65209458803974
- Mean Absolute Percentage Error: 33.109951301919%



# Random Forest Regression Model

- Mean Squared Error: 1132.4634701623736
- Root Mean Squared Error: 33.65209458803974
- Mean Absolute Percentage Error: 33.109951301919%



# DATA SELECTION

Add more Features


Features such as job experience, data related experience and education qualifications

# CLASSIFICATION

Using Classification instead of Regression

By creating different bins for salary.

# CHALLENGES

- Data Collection: Choosing the right data
  - Model: Choosing the right model
- 
- The background is a solid light green color. It features several abstract, organic shapes in a slightly darker shade of green. These shapes are located in the bottom left, bottom right, and middle right areas of the slide, creating a modern, minimalist aesthetic.



Thank you!