Salary Calculator for Data Science and Analytics

Presented by: Sanjay Dabra, Pranay Nuvvala, Kai Murata, Richard Kim

Agenda

- Introduction
- Purpose
- Data Preparation
- Machine Learning, Training, and Testing
- Demo of App

Introduction/Purpose

PayDay

There are many people employed in Data Science and Analytics, and many more candidates looking to enter.

Current employees and candidates have concerns about compensation.

Create an application to quickly aid candidates and current employees for salary negotiations, check to see if they are appropriately compensated, and future growth potential.

How compensation is affected by experience, field of employment, and annual change.

Data Source

Stack OverFlow Annual Salary Survey

2016 to 2022

Focused on Year of the Survey, Experience, Field of Employment, and Salary

Data Preparation

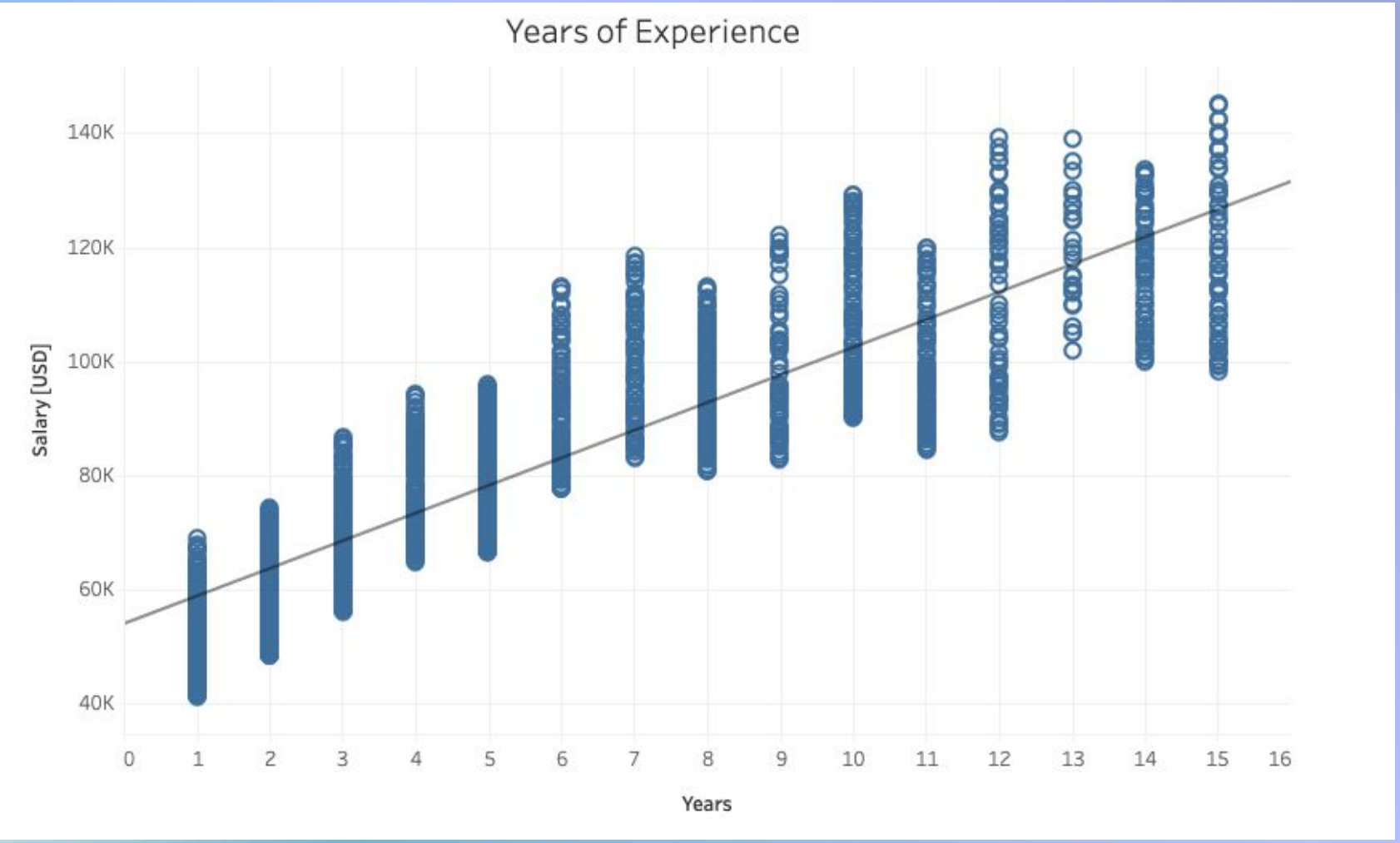
Data Filtering

- Filter down salary to $500,000 and Experience Year <= 15 years
- Filter salaries based off years of experience and salary percentile
  - For each experience year, the middle 50% was used
- All filtering was performed for higher machine learning classification accuracy

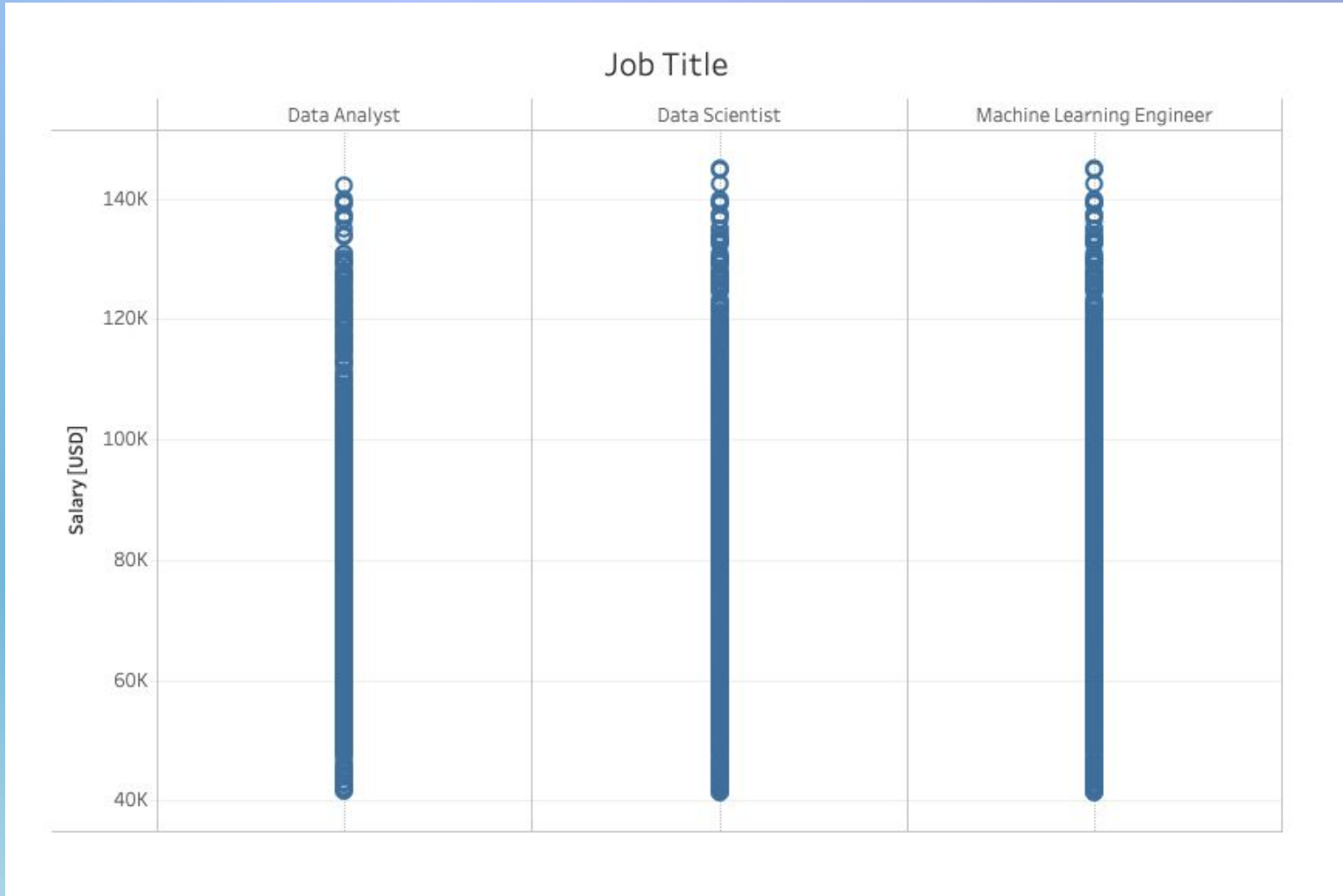
Model Analysis

	Years of Experience	Job Title	Year (Date)	Combined
Training Score	77.5%	0.012%	0.285%	77.8%
Testing Score	76.1%	0.112%	0.652%	76.1%
Formula	$y = 4999x_1 + 53533$	$y = -298x_2 + 82921$	$y = 777x_3 - 1486966$	$y = 4999x_1 + 77x_2 + 765x_3 - 1491542$

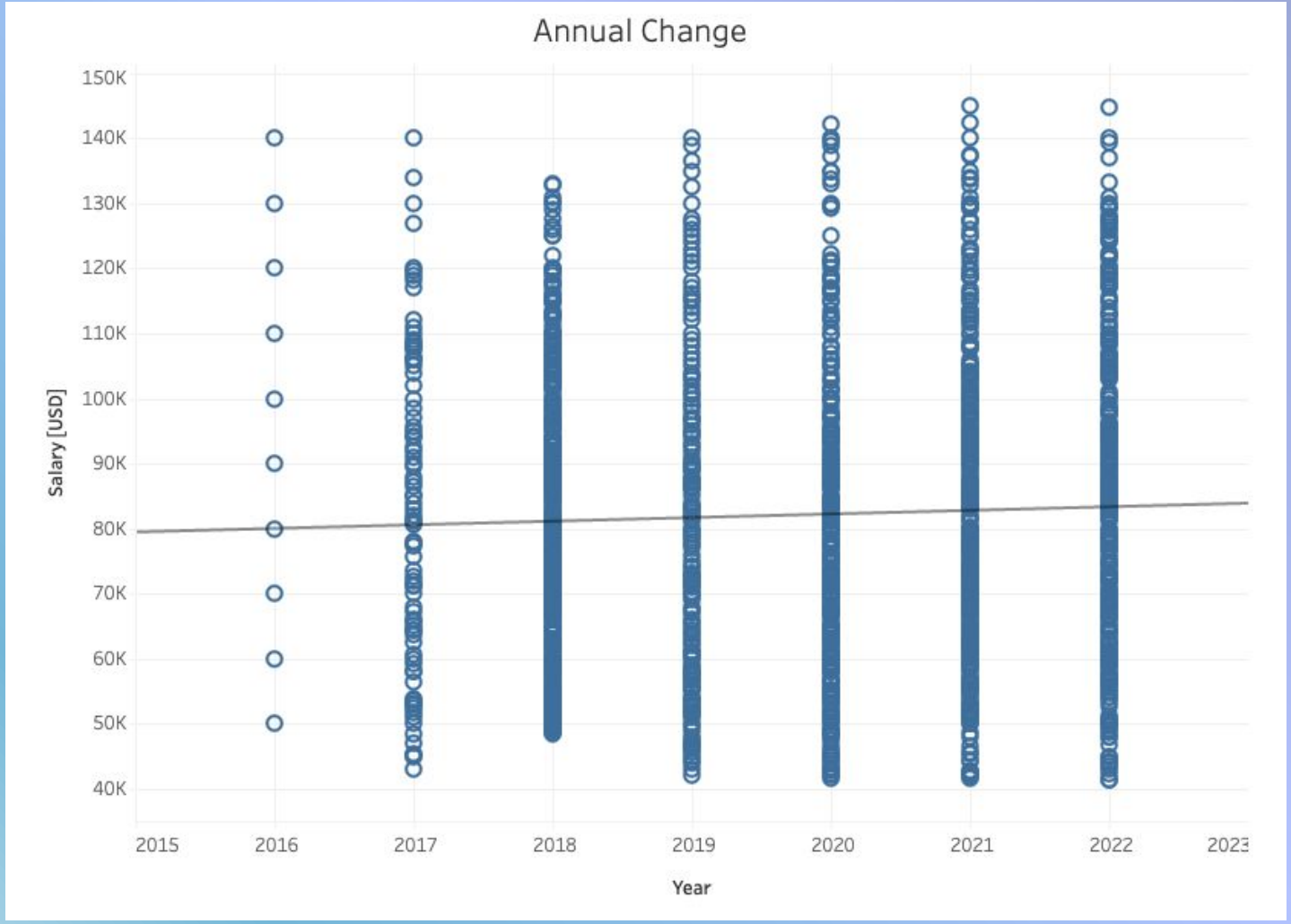
	Years of Experience	Job Title	Year (Date)	Combined
Training Score	77.5%	0.012%	0.285%	77.8%
Testing Score	76.1%	0.112%	0.652%	76.1%
Formula	$y = 4999x_1 + 53533$	$y = -298x_2 + 82921$	$y = 777x_3 - 1486966$	$y = 4999x_1 + 77x_2 + 765x_3 - 1491542$



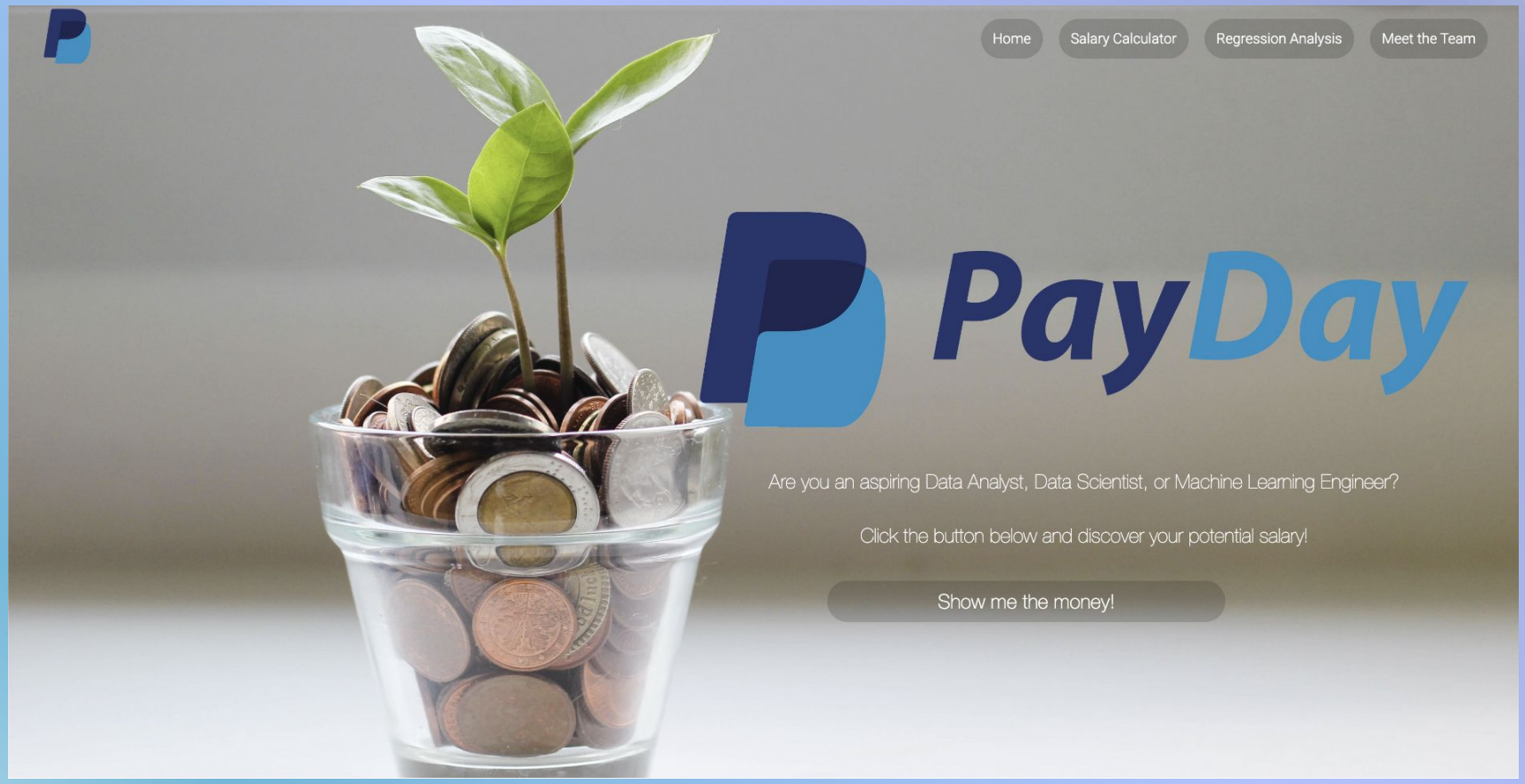
	Years of Experience	Job Title	Year (Date)	Combined
Training Score	77.5%	0.012%	0.285%	77.8%
Testing Score	76.1%	0.112%	0.652%	76.1%
Formula	$y = 4999x_1 + 53533$	$y = -298x_2 + 82921$	$y = 777x_3 - 1486966$	$y = 4999x_1 + 77x_2 + 765x_3 - 1491542$



	Years of Experience	Job Title	Year (Date)	Combined
Training Score	77.5%	0.012%	0.285%	77.8%
Testing Score	76.1%	0.112%	0.652%	76.1%
Formula	$y = 4999x_1 + 53533$	$y = -298x_2 + 82921$	$y = 777x_3 - 1486966$	$y = 4999x_1 + 77x_2 + 765x_3 - 1491542$



Deployed Model



Conclusion

- From the machine learning analysis, years of experience was shown to have the most impact on salary
  - Job title and year was shown to have less impact on salary

- Future Analysis
  - Using salary information from various data sources for a better understanding of salary ranges
  - Using additional features such as company size, location, and demographic information (age, race, sex, etc)

Sources

- https://insights.stackoverflow.com/survey
- https://html5up.net/

The End

- https://insights.stackoverflow.com/survey
- https://html5up.net/