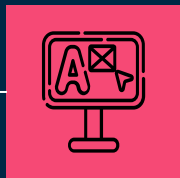


Predicting a Users Operating System

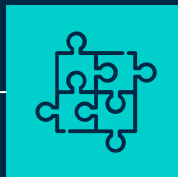
Rachel Dilley

OUTLINE



01

Introduction



02

Data & Model
Selection



03

Model Tuning &
Final Model



04


Next Steps

INTRODUCTION

01

A collection of small squares in various colors (cyan, pink, orange) scattered in the top right corner of the slide.

What is an Operating System?

- Manages a computer's memory and processes, as well as all of its software and hardware
 - **Windows, MacOS, Linux**
- 
- A small cluster of squares in cyan and orange colors located in the bottom left corner of the slide.

Why is it Important to Know What Operating System a User Has?

- Some softwares and applications are **specific** to some operating systems
- Useful for **software companies** looking to market to users with certain operating systems

DATA & MODEL SELECTION

02

DATA

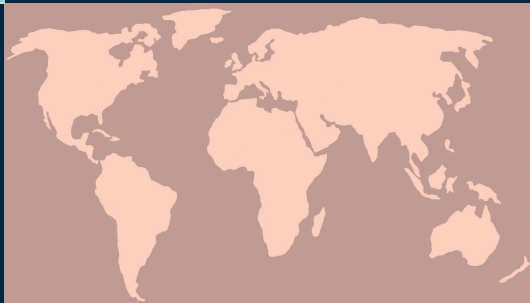


STACK OVERFLOW SURVEY DATA

- 2019 & 2020
- Over 75,000 records

COUNTRY DATASET

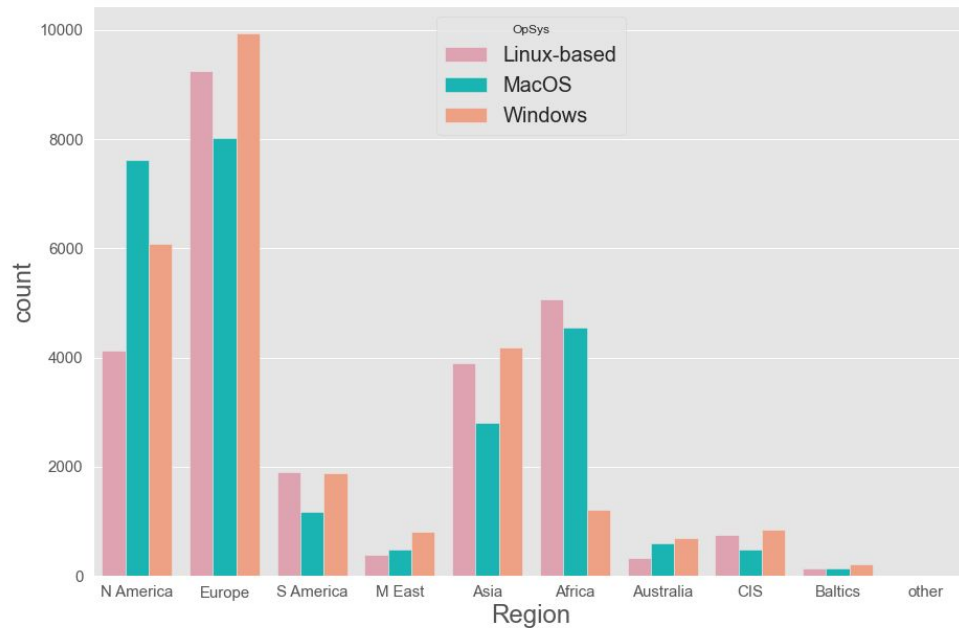
- For region data



FEATURES

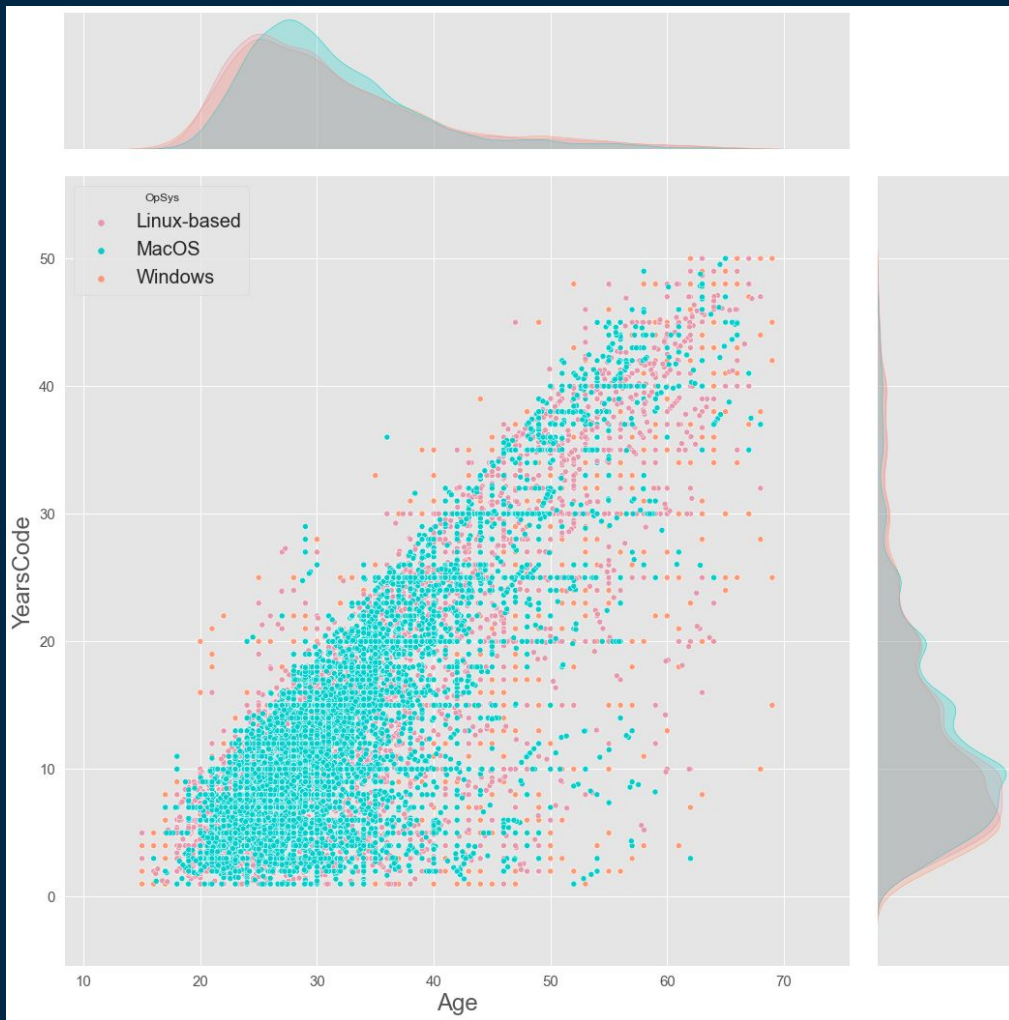
- Age
- Region
- Gender
- Education level
- Undergrad major
- Age user started coding
- Years of coding experience
- Years of professional coding experience
- Number of databases used
- Top coding languages used
- Top developer types

REGION



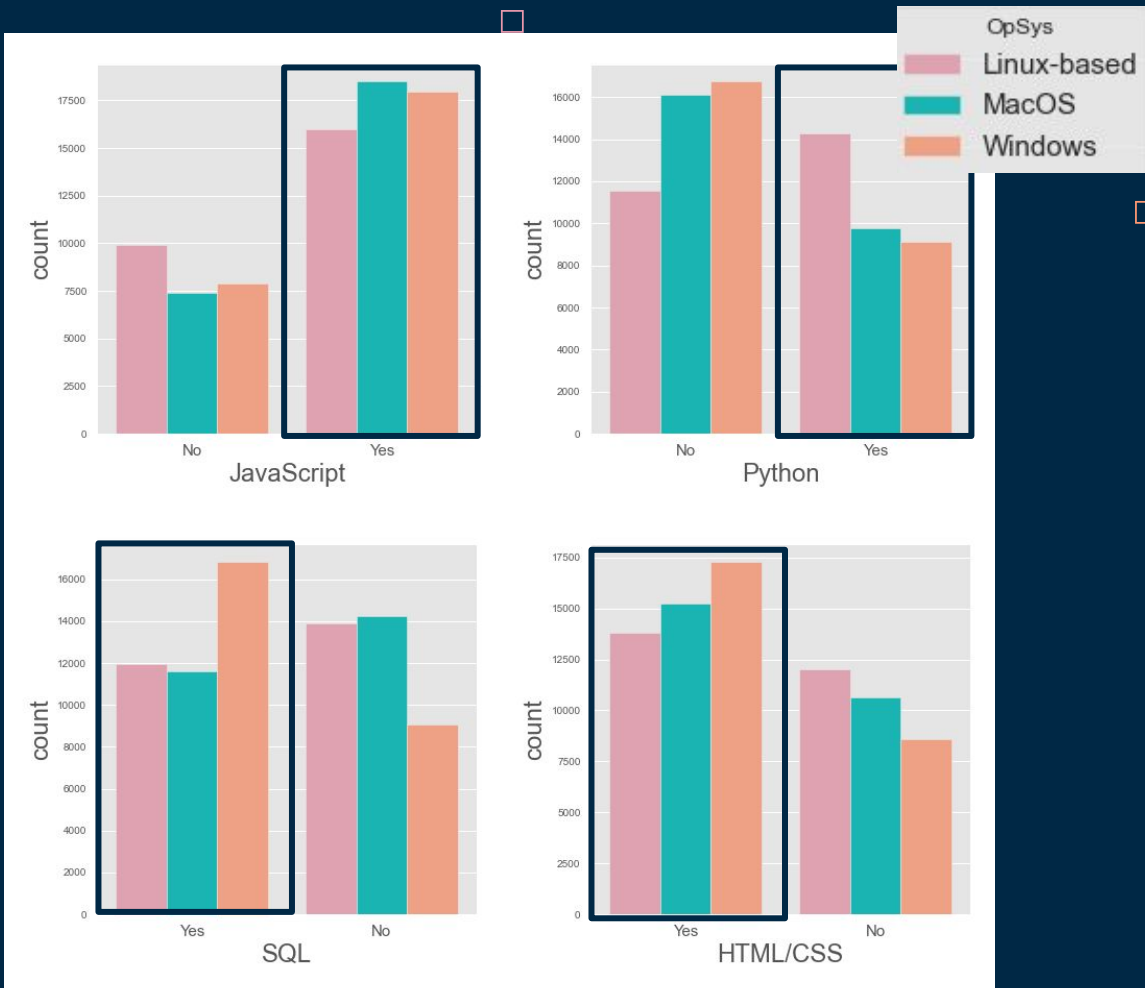
AGE & YEARS OF CODING EXPERIENCE

- **Windows**
 - Younger,
 - Less experienced
- **MacOS**
 - Ages 25-40
 - 10-20 years experience
- **Linux**
 - Older,
 - More experienced



TOP CODING LANGUAGES

- **Java/JavaScript** primarily **MacOS**
- **Python** primarily **Linux**
- **SQL** and **HTML/CSS** primarily **Windows**



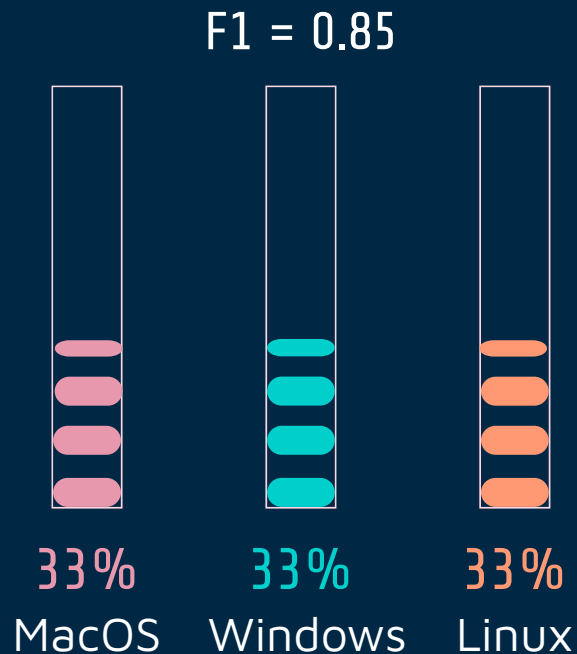
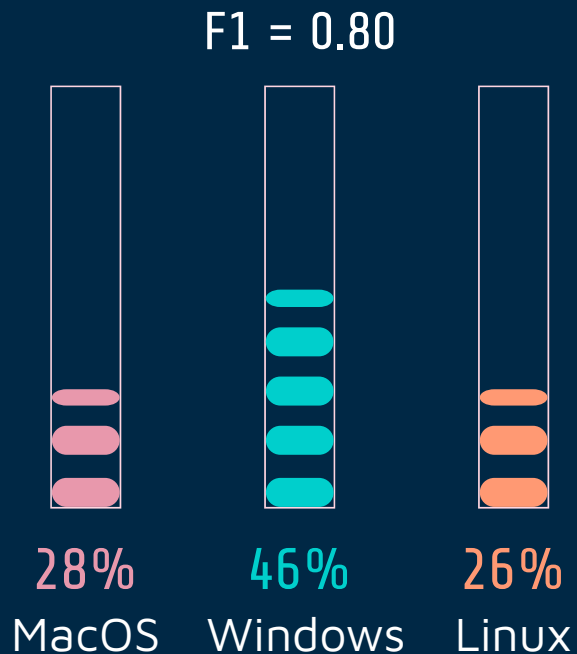
MODEL SELECTION

	Hyperparameter(s)	F1 Score
LOGISTIC	$C = 1$	0.50
KNN	$K = 5$	0.52
DECISION TREE	Depth = 14	0.52
RANDOM FOREST	Depth = 20	0.70
XGBoost		0.80

MODEL TUNING & FINAL MODEL

03

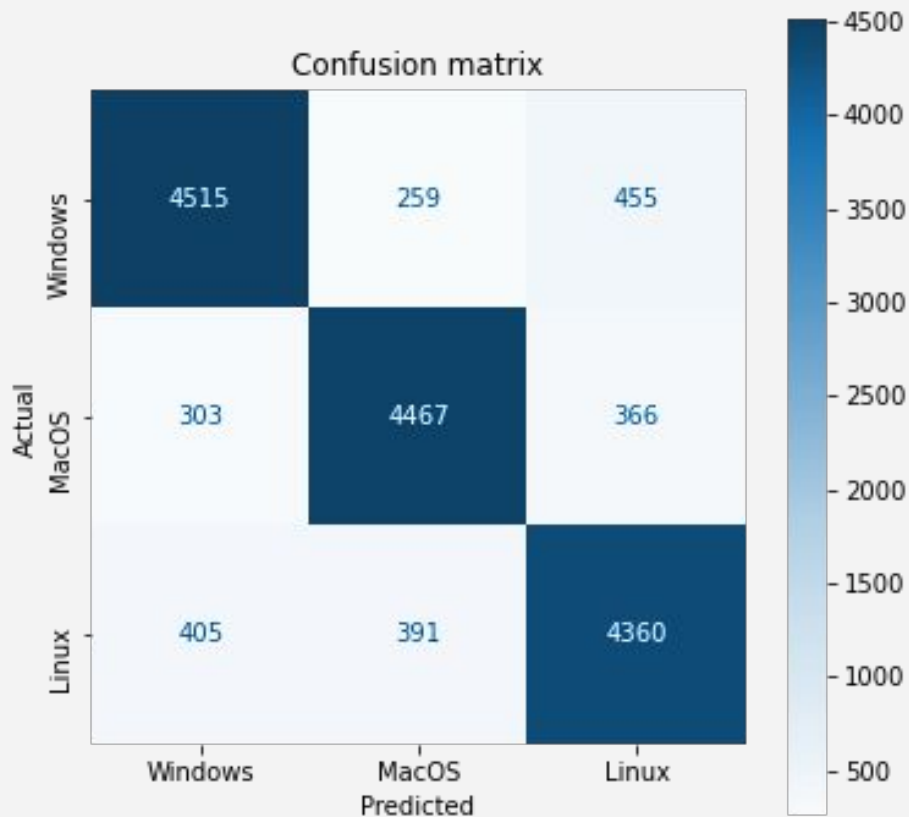
OVERSAMPLING

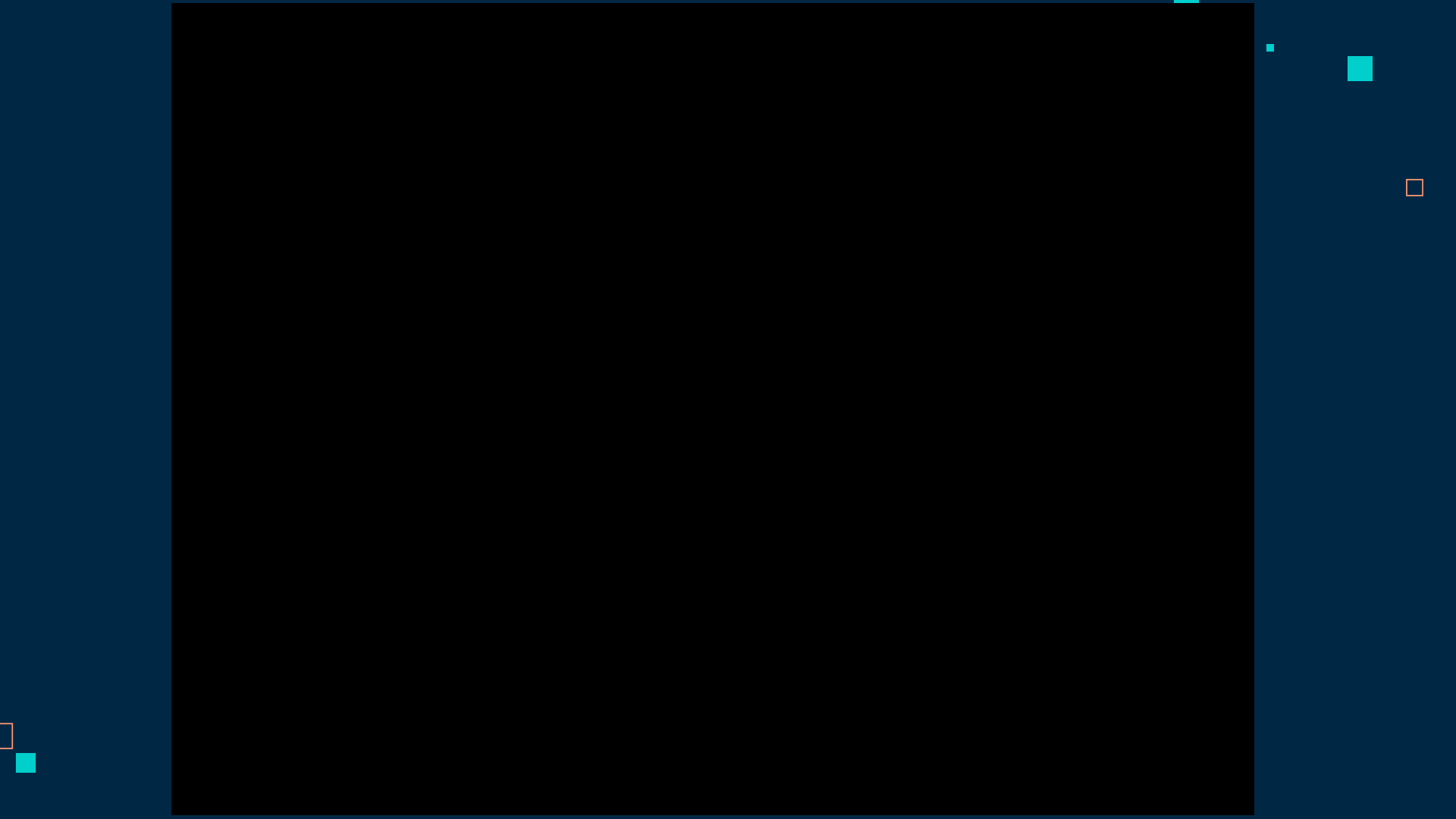


FINAL MODEL USING XGBOOST

Precision, Recall,
and F1 Score:

0.86





NEXT STEPS

04

NEXT STEPS

- Include more features
 - More coding languages
 - More developer types
- Continue tuning hyperparameters
- Expand the scope
 - Include non-programmer specific data

The background is a dark blue gradient. It is decorated with various geometric elements: thin white vertical lines of varying lengths, small squares in light blue, pink, and orange, and larger solid squares in light blue and pink. The word "Questions?" is centered in a large, orange, sans-serif font.

Questions?

FEATURE IMPORTANCE

