# Singapore Public Residential Properties (HDB)

Presented by: Andre Tan

GA GENERAL ASSEMBLY

HOUSING & DEVELOPMENT BOARD

# TABLE OF CONTENTS

## Introduction

- **80% of Singaporeans live in HDB**
- **Housing is the probably the largest purchase in one's lifetime**

## Problem Statement

- **The objective of this project is to design a regression model to predict resale prices of HDB flats, using past 5 years resale HDB data from data.gov.sg.**

## HDB

## Homeowners

## URA

## Real Estate Agents

# DATA & METRICS

## Data

- **Jan 2017 - Apr 2022**
- **125,360 data points**


Data.gov.sg

## Metrics

- **MAE**
- **RMSE**
- **R² score**

| **Highest Floor** | **Highest resale price** | **Highest $PSF** |
|:---:|:---:|:---:|
| 51 | 1.388m | $1,226 |

| **Largest Size** | **# of units > 1m** | **Cheapest** |
|:---:|:---:|:---:|
| 2680 sqft | 557 | $194 psf<br>$140k |

## Singapore — by postal district

| LOCALITIES | DISTRICTS |
|---|---|
| City & Southwest | 1 to 8 |
| Orchard/Tanglin/Holland | 9 and 10 |
| Newton/Bukit Timah/Clementi | 11 and 21 |
| Balestier/MacPherson/Geylang | 12 to 14 |
| East Coast | 15 and 16 |
| Changi/Pasir Ris | 17 and 18 |
| Serangoon/Thomson | 19 and 20 |
| West | 22 to 24 |
| North | 25 to 28 |

No. of sale transactions above 1m, by town (2017 - 2022)
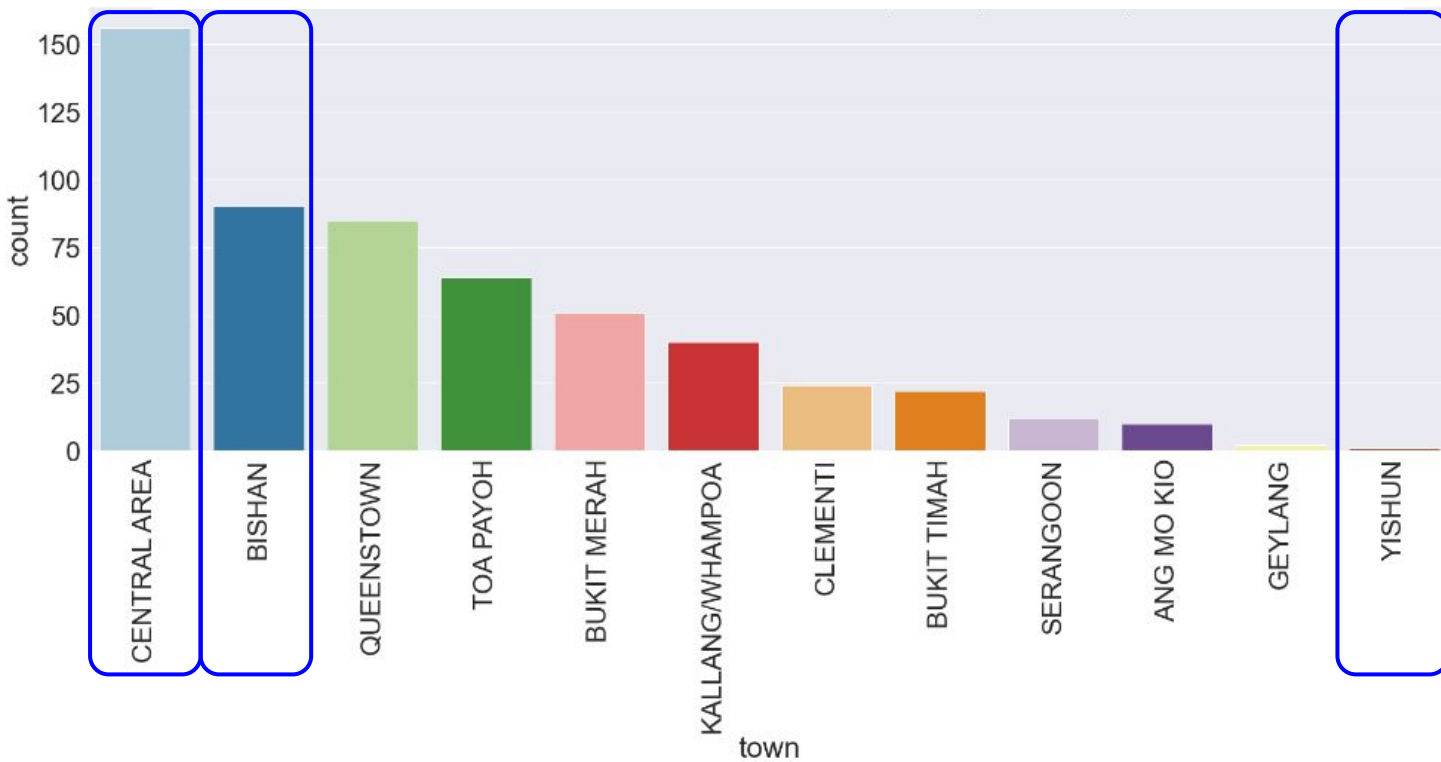
No. of sale transactions per district, year on year comparison (2017 - 2022)
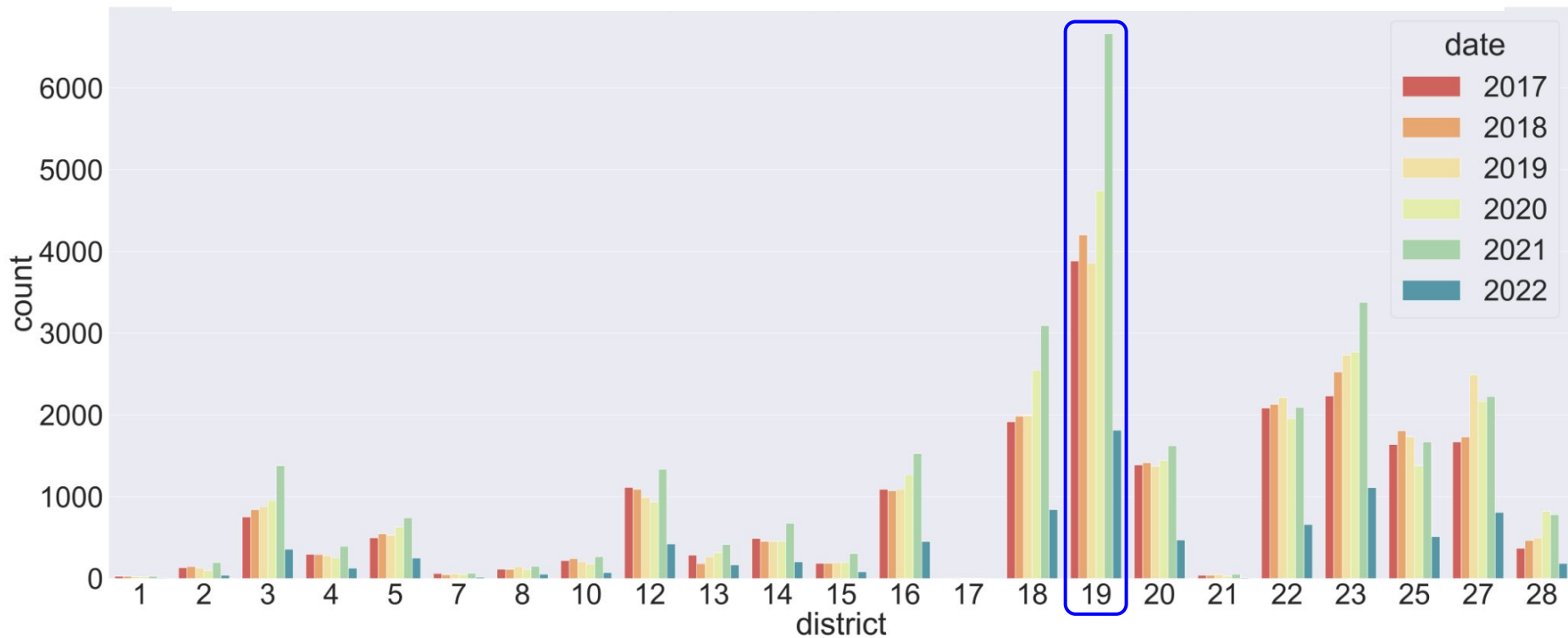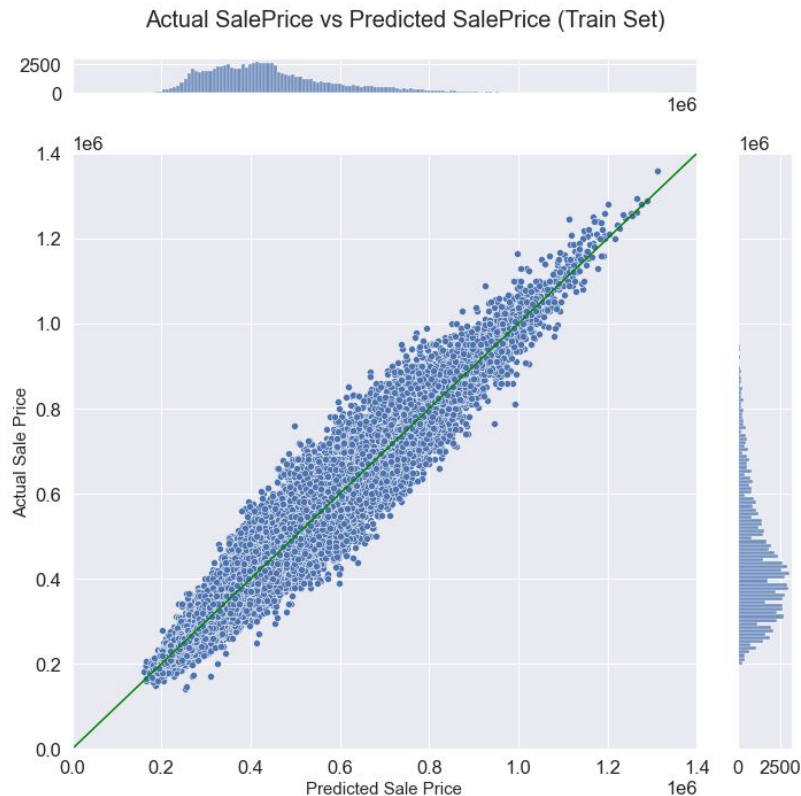
# MODELING

| MODEL | MAE | RMSE | R2 |
|---|---|---|---|
| GradientBoostingRegressor | 44358.877 | 58879.978 | 0.857 |
| RandomForestRegressor | 51356.424 | 67362.893 | 0.813 |
| LinearRegression | 56402.681 | 72854.322 | 0.782 |
| KNeighborsRegressor | 60598.267 | 76563.529 | 0.759 |
| AdaBoostRegressor | 71504.518 | 105946.896 | 0.538 |

## MODELING

- **Gradient Boosting Regressor**
  - **MAE:     44,358.877**
  - **RMSE:   58,879.978**
  - **R²:         0.857**

- **Model explains 85.7% of variability in sale price**



Actual SalePrice vs Predicted SalePrice (Train Set)

Gradient Boosting Regressor - top 10 feature importance
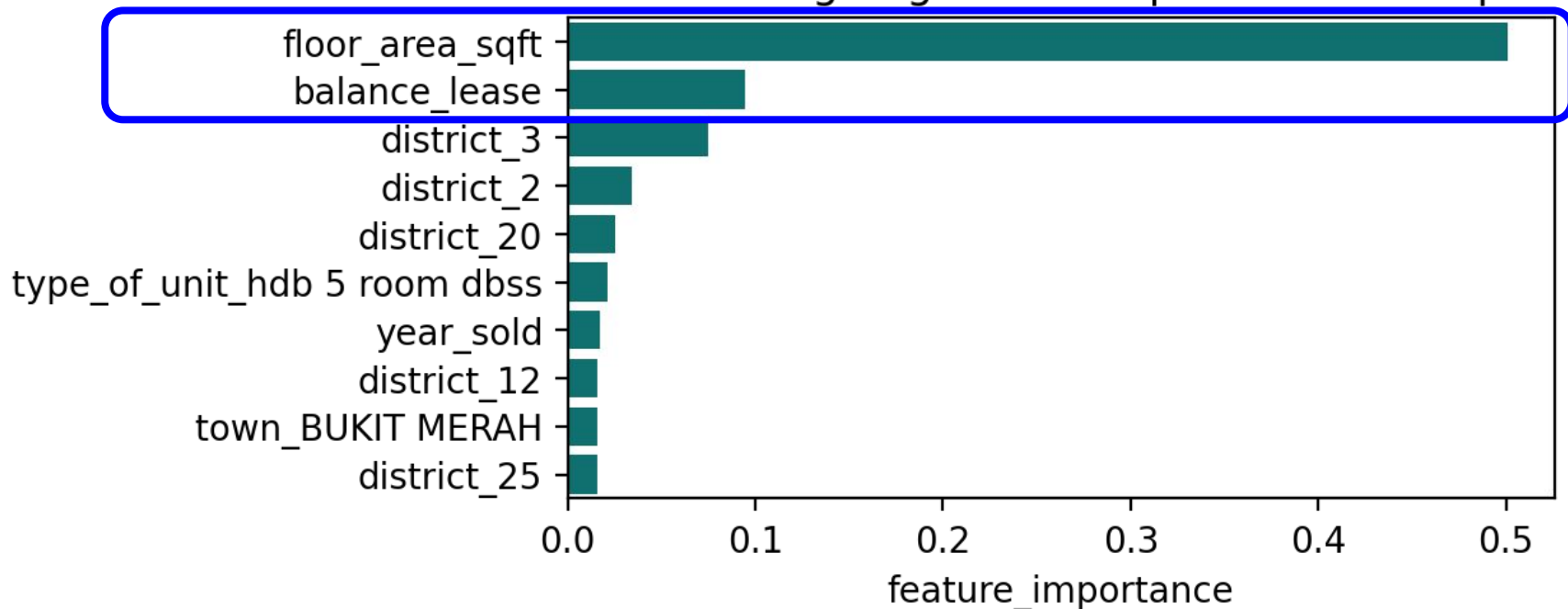
# CONCLUSION & RECOMMENDATIONS

- **Top 3 features**
  - **Floor area**
  - **Balance lease**
  - **Location**

# LIMITATIONS, MOVING FORWARD

## Limitations
- **Need to retrain model**
- **Only applicable to HDB**

## What's next?
- **Deploy to dashboard**
- **Consider to impute other features:**
  - Proximity to amenities, transportation, etc
  - Renovation, theme, quality of build
- **Include private property data**

# Massive thanks to:

- **Teaching Team - Divya, Shao Quan, Ben**

- **Career Coach - Samantha**

- **DSIF4 classmates**

- **Inspiration - Yuan Tie Ming**

- **Pillar of support - MY WIFE!**

# Q&A