

0000

PRESNTATION  
ARNOLD WANG

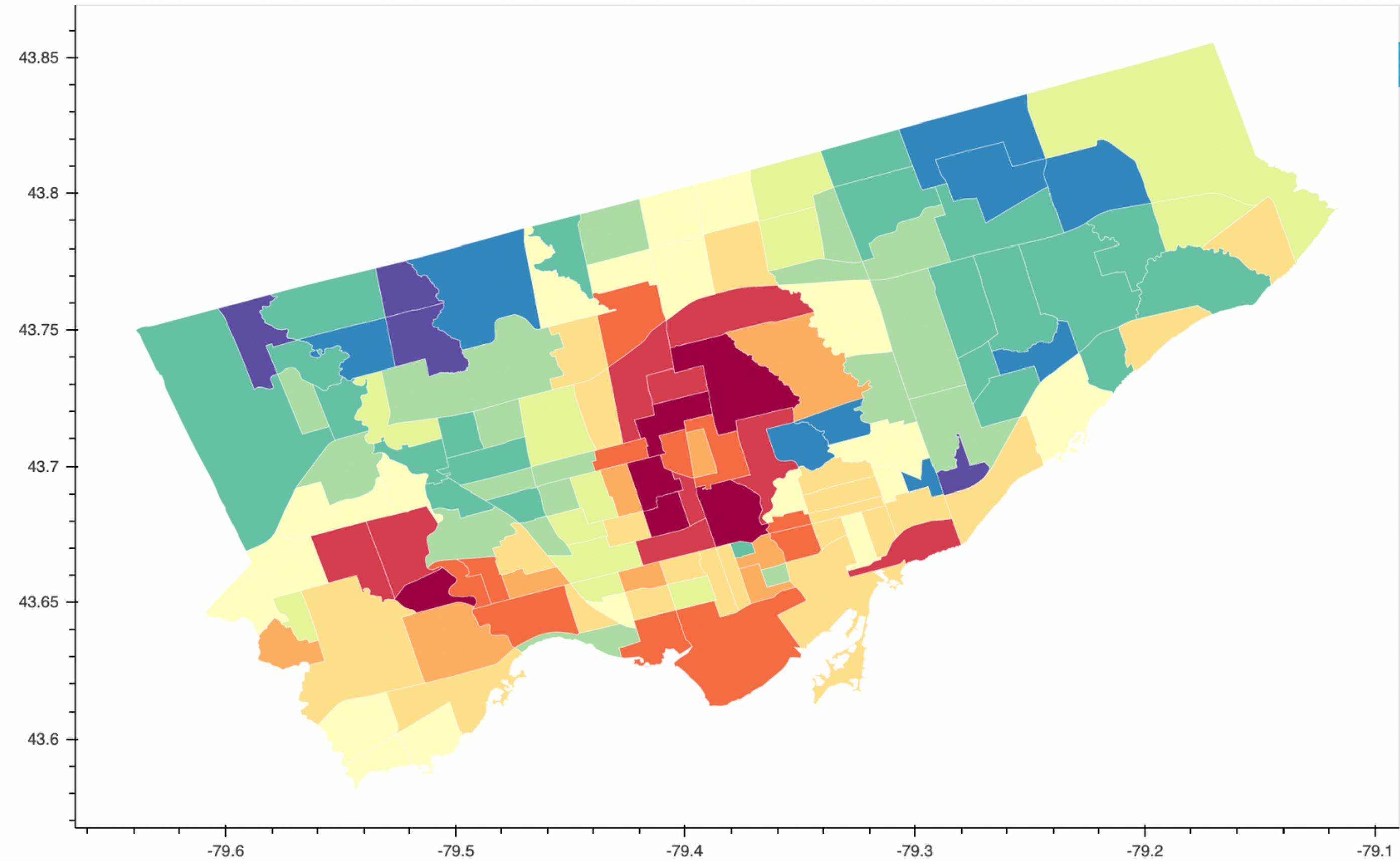
# REAL ESTATE ESTATE MODEL

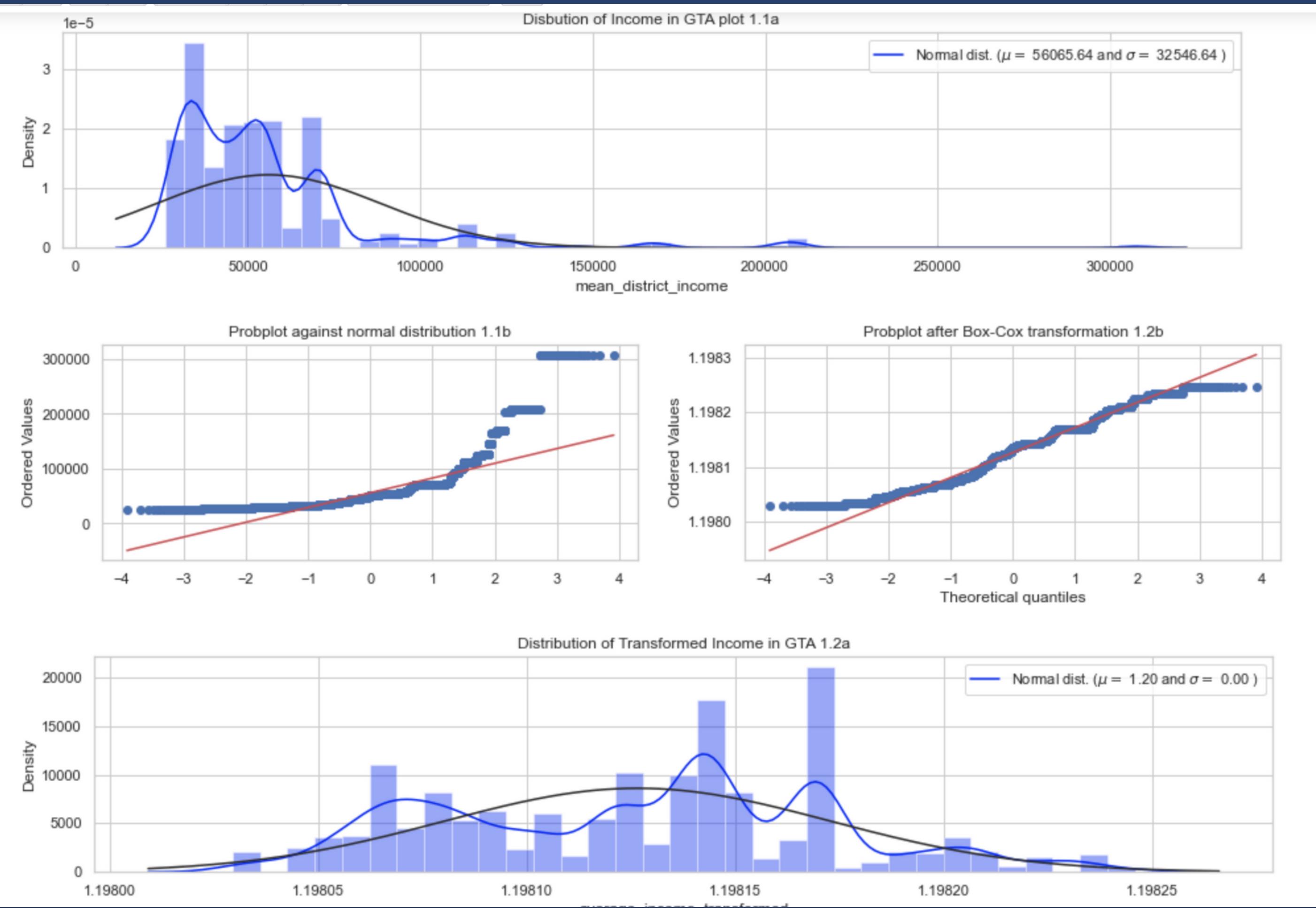
0000



# HOW MUCH DO PEOPLE MAKE AND WHERE

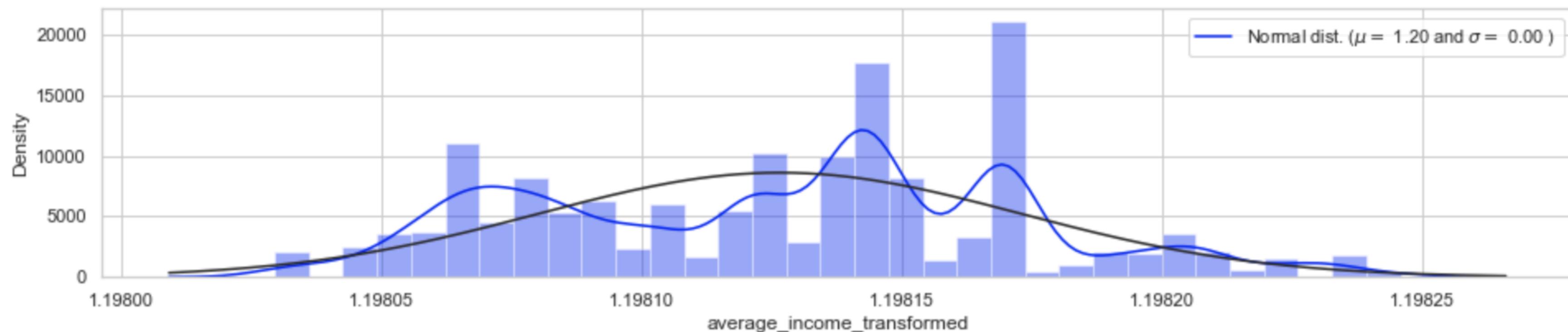
GTA Neighbourhood Income



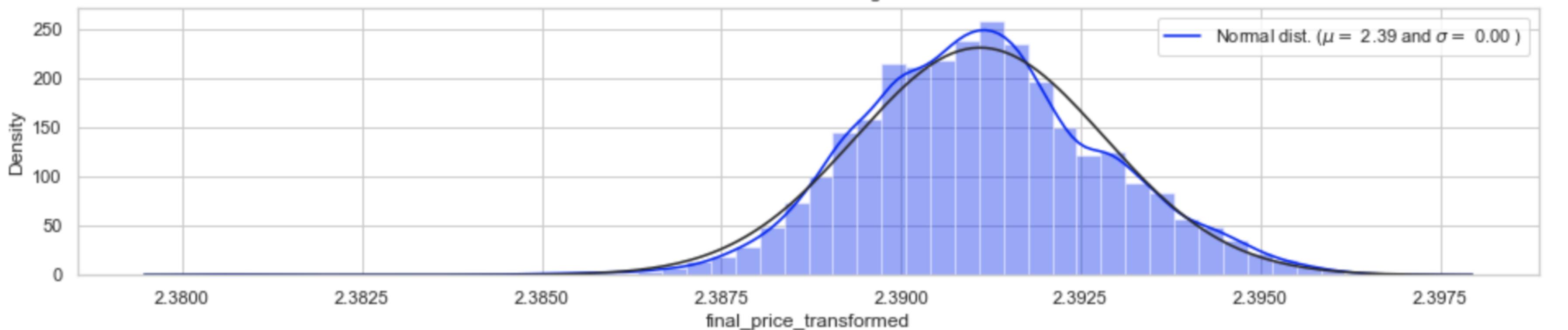




Distribution of Transformed Income in GTA 1.2a



Distribution of Transformed Housing Prices in GTA 2.2a

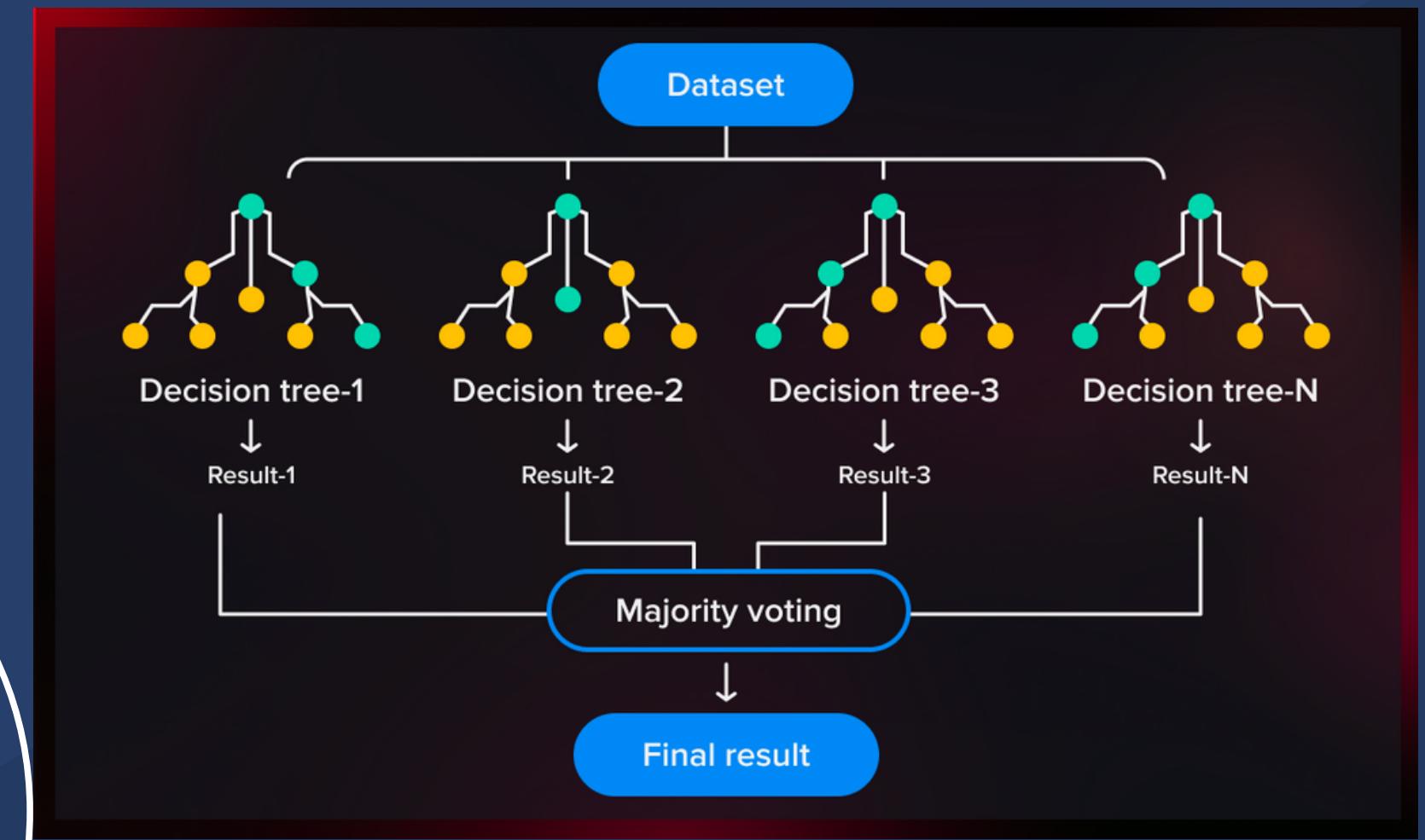






# KGBoost

"eXtreme Gradient Boosting", and it is designed to be fast and efficient, making it a popular choice for large-scale data analysis.



## Random Forest

large number of randomly-generated decision trees to make predictions.

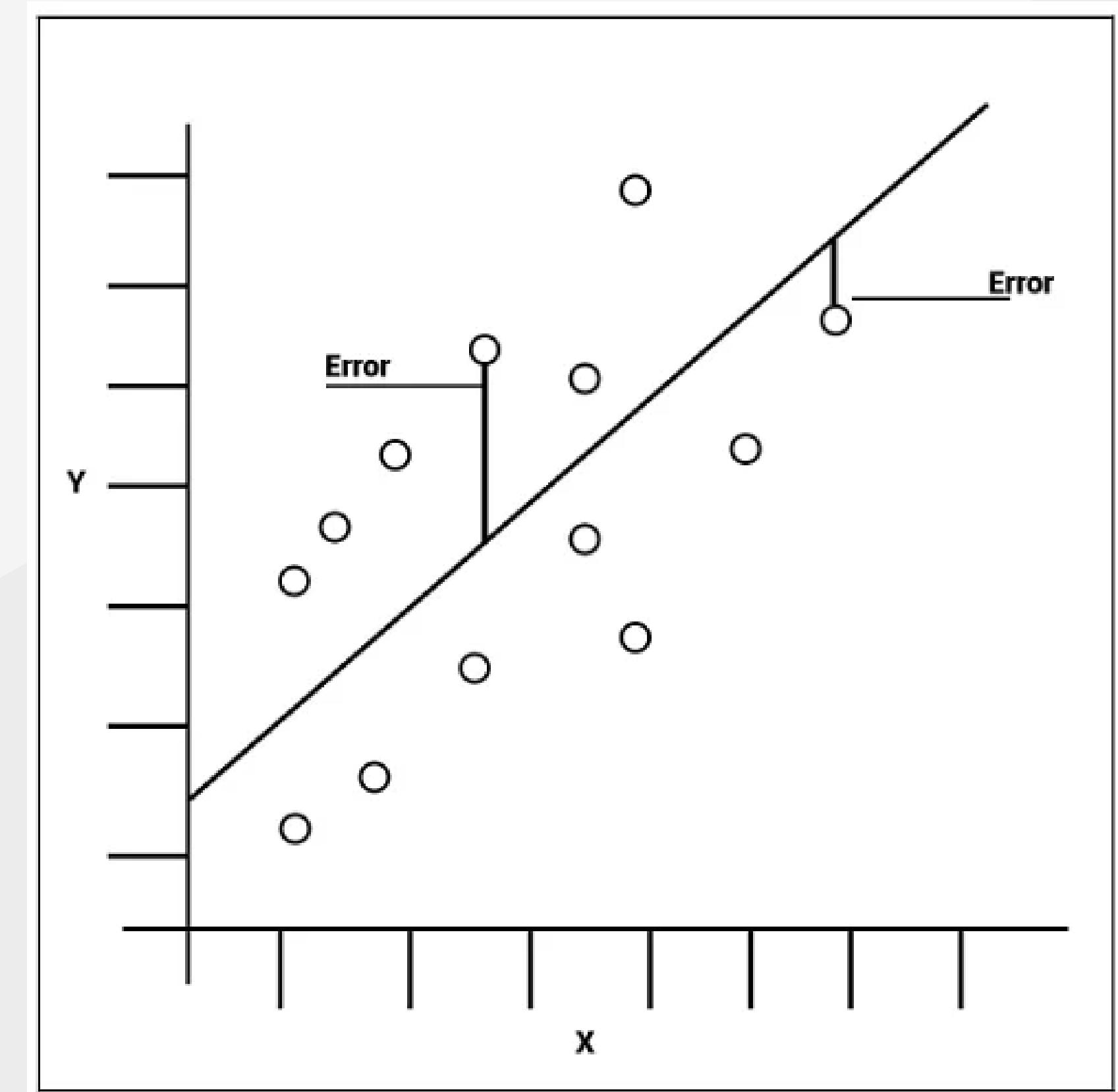
o o o

# RESULTS

- No confusion matrix :)
- Use Mean Squared Error
- Use Mean Absolute Error

MAE ~40 000 (predictions were off by around that)

o o o



# FUTURE PLANS

- More Data!
- More Models!
- Include dates for time series analysis and seasonality feature importance
- When measuring accuracy, focus on not only mae and rmse but other factors as well
- Use hyperopt and bayesian search methods

