

# Kaggle Project Presentation

Keqi Yu

- **What you did right with the analysis and where you went wrong.**

Model: Linear Regression, Decision Tree, Boosting Model(gbm & xgboost), Random Forest Model

Model	RMSE on scoring data in R	Score on Kaggle	Notes
Model 1: Linear Regression	142.293	106.32416	
Model 2: Decision Tree	133.1909	86.37468	
Model 3: gbm model	127.2261	87.59981	
Model 4: xgboost model	140.4818	72.32	Final Kaggle result RMSE
Model 5: Random Forest	138.0336	71.61806	The lowest score

**Best Model with lowest RMSE:  
Xgboost score 72.32 in Kaggle**

**The most impactful 6 variables:**  
Accommodates,  
Bedrooms,  
Bathrooms,  
calculated\_host\_listings\_count\_entire\_homes,  
calculated\_host\_listings\_count\_private\_rooms,  
Neighbourhood\_group\_cleashed.

- **If you had to do it over, what you would do different.**

I will spend more time on selecting the variables, also try more models in order to find the best model