

Paris Real Estate Property Predictions



Deztany Jackson
Data Science Flex
Phase 3 Project
February 24, 2023

Overview

Classification models will be built to best predict "Luxury" property for a Paris Real Estate investment agency to support leasing to potential clients

Model: Binary Classification

Classification Values: Basic, Luxury

Agenda

- Business & Data Understanding
- Modeling
- Results
- Limitations
- Recommendations
- Next Steps

Business and Data Understanding

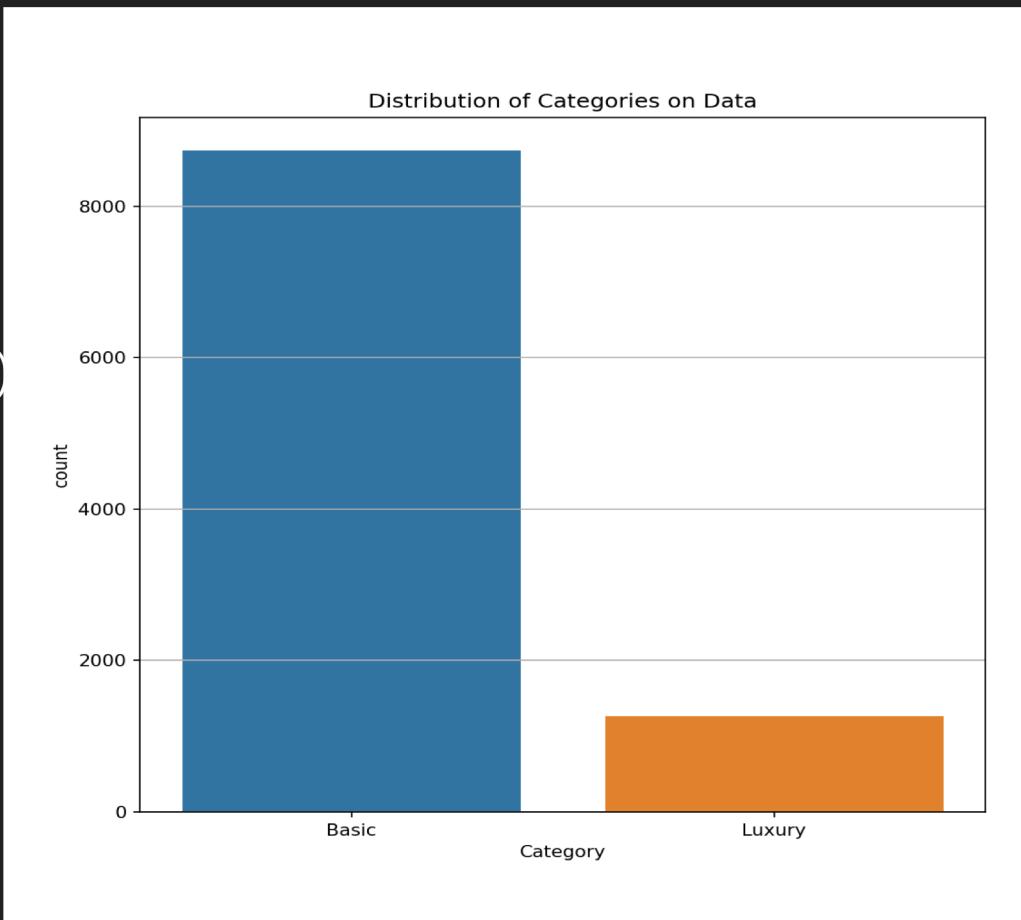
- **Stakeholder:** Paris Real Estate Agency
- **Use case:** Predict “Luxury” Property Data
- **Dataset:** 10000 Entries; 18 Features/Attributes
- **Target Variable:** Category [“Basic”, “Luxury”]

Target Attribute	Description
Category	Classification of Real Estate (Basic or Luxury)

Importance Level	Top Important Features
1	Has_Yard
2	Has_Pool
3	Num_of_Guest_Rooms
4	Num_of_Rooms
5	Building_Size(m)

Distribution of Categorical Data

- Basic Instances: 8735/10000 (87%)
- Luxury Instances: 1265/10000 (13%)



Modeling

Iteratively produce models. As new information was learned new models, parameters and transformation techniques were applied.

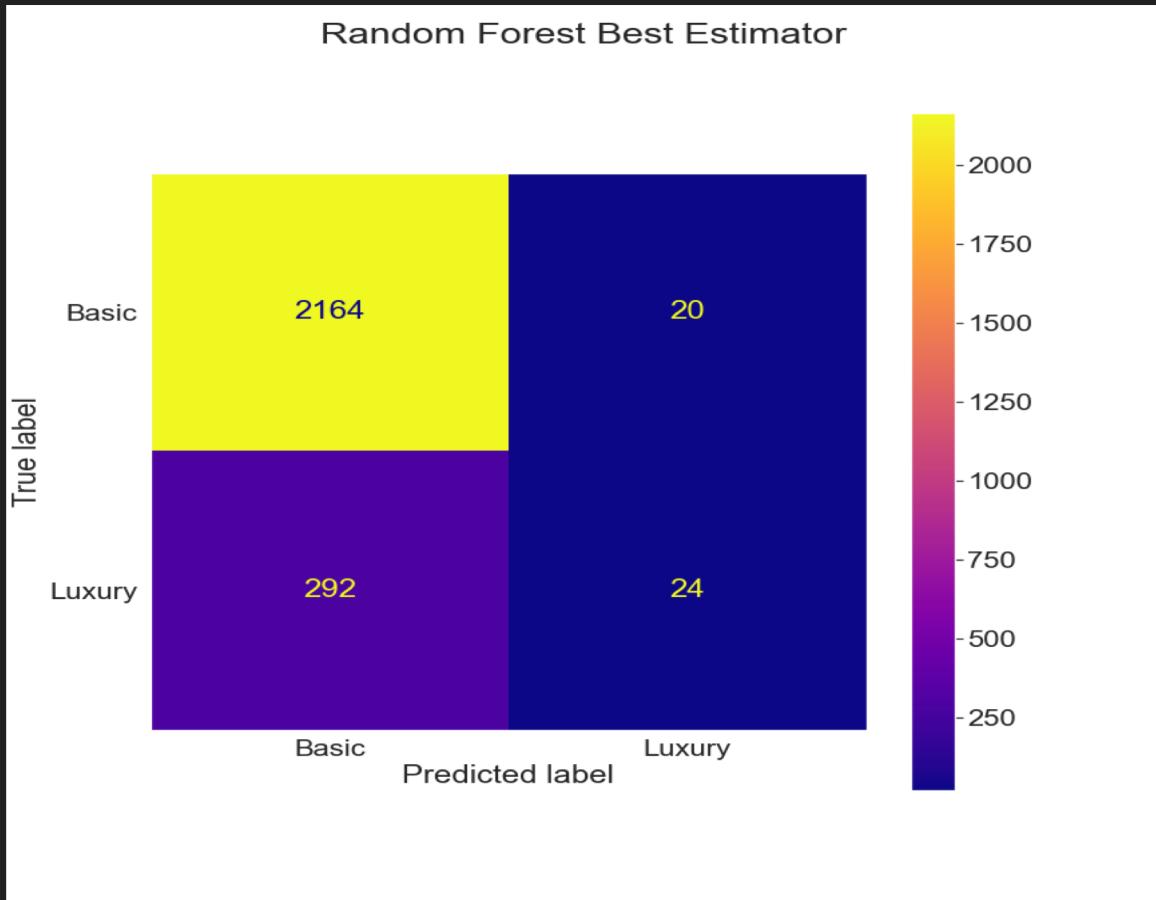
1. Dummy Model
2. Simple Model: Logistic Regression, Decision Tree and Random Forrest
3. Standardizations & Balancing Data
4. Best Model from optimization of Parameters

Evaluation

- Best Model:
 - Standardize Data
 - Synthetic Data
 - Random Forest Classifier (w/ 10 Decision Trees)
- The Final metrics:
 - Precision Score (Correct Predictions) : .52 (Better than Random Guessing)
 - F1 Score(Harmonic Mean): .13
- Most Important Features (Ref Slide 4.):
 - Yard and Pool

Best Model's Prediction

- Accurate “Luxury” Prediction: 24
- Incorrect “Luxury” Prediction: 20
- Accurate “Basic” Prediction: 2164
- Incorrect “Basic” Prediction: 292



Conclusion: Limitations

- Imbalanced data between "Basic" and “Luxury” properties
- Used synthetic data
- Limited computing resources
- Limited knowledge of stakeholders requirements
- No knowledge on the condition of property.

Conclusion: Recommendations

- Recommend using the "best" model for a small list of "Luxury" property listings
- Recommend looking any property that has a yard, pool and increased guest rooms.
- Recommend not using "best" model for initial, not all searches
- Recommend identifying criteria requirements for top features
- Recommend new modeling with more than two classification values

Conclusion: Next Steps

- **Refine and Iterate more models for higher precision and accuracy scores**
- **Gather more real data on “Luxury” property**
- **Increase consultation with the data scientist/analyst to improve our domain knowledge and more specific feature criteria**

Thank You!

Email: dmvinedata@gmail.com

GitHub: [@dmvinedata](https://github.com/dmvinedata)

LinkedIn: [linkedin.com/in/deztanyjackson/](https://www.linkedin.com/in/deztanyjackson/)