This his the project BnB (Black and Blue) of the French Mexican Collab for the Data Analytics Class.

Our work consists in analyzing the Airbnb listings of New York City in order to allow our parents and friends to make the wisest choice when they come to Columbia for our graduation.

Our project is organized as follow:

1. Files which contains every file we used and created for our work
2. Data Preprocessing folder which contains every file that is necessary to scrap Airbnb website and clean our dataset in order to perform our analysis
3. Models which contains the ML models
4. Sentiment Analysis where we performed the analysis of reviews of Harlem, Morningside Heights and Upper West Side
5. Visuals which contains some nice visualizations of our datasset
6. PPT which contains our presentation

These are the parts we would like to highlight:

1. Scrap the Airbnb website was difficult. We scrapped the reviews and the ratings of over 5000 listings (when the listing were still available)
2. Build Several machine learning models (Linear Regression, K-Nearest Neighbor, Random Forest, Gradient Boosting) and tune them in order to achieve better performances (Add features, limit number of features to avoid overfitting, get rid of outlier values that weren't relevant for our purpose, indeed we don't want our parents and friends to pay too much, use grid search algorithm to fine tune the parameters)
3. Perform a sentiment analysis over the three neighborhoods using Vader, NRC, display a wordcloud ans a word summarization to understand what people talk the most about and construct a LDA model to see the important topics (finally limited just to one topic, since it was always the same words and patterns, eventually this model was not very relevant)
4. Add a visualization file using folium library to nicely display some features of our dataset on NYC map

Thanks to the new skills we acquired in this class, we have been able to perform these analysis and to work on the skills we acquired during the semester.