**Airbnb**

**Abstract:**

**The goal of this project was to use Airbnb reviews for NLP and topic modeling. I worked with data provided by inside Airbnb. After topic modeling, I did some sentiment analysis but it did't give me a good result so I can do a classification models om them.**

**Design:**

**This project addresses two applications that can be used by Airbnb. with topic modeling, Airbnb can gain insight on what users love and hate. They can work with the hosts and using the findings, they can improve customer experience from multiple angles. Also, sentiment analysis it supposed to help Airbnb to identify negative and positive reviews and know what the topics is they are talking about in the negative reviews so they can handle it.**

**Data:**

**The dataset I downloaded contains about 84000 reviews of listings in Seattle cite from Inside Airbnb.**

**Algorithms:**

**Natural Language Processing was used to preprocess the Airbnb reviews. CountVictorizer and TF-IDF Vectorizer were used to turn review documents into sparse matrices for further use in modeling. NMF, LSA, LDA with grid search, and CoreX with no anchors and with weak and strong anchors were used for topic modeling before settling on CoreX with weak anchors because it produced the most distinct topics.**

**Topics are:**

* **Rooms**
* **Location**
* **Host Communication**
* **Cleanliness**
* **Also we used Vader tool for sentiment analysis**

**Tools:**

**Numpy and Pandas for data manipulation Scikit-learn, Gensim, and CoreX for topic modeling Matplotlib, Seaborn, and pyLDAvis for data visualization NLTK for NLP Vader for Sentiment analysis.**