MIDTERM PROJECT REPORT

WEB ANALYTICS

SCRAPE DATA FROM FREELANCER.com using Python
SENTIMENT ANALYSIS TO IDENTIFY POSITIVE OR NEGATIVE REVIEWS
OBSERVE DIFFERENT CORRELATIONS

MALHAR INAMDAR

Introduction:

Freelancer.com is a website which provides freelancers for thousands of jobs from web design, mobile app development, product design to manufacturing

Goals and objetives:

1. To scrape the data for top 200 user profiles from freelancer.com
2. To identify whether each user has positive / negative reviews
3. To observe different correlations by using Visualizations

The main attributes which we collected from each user for all skills are :

1. Username
2. Skill Rating
3. Country
4. Amount charged per hour

The techniques which we used for the project are

1. Data collection/cleaning/Preparing data files using Beautifulsoup library in python
2. Text, Sentiment Analysis
3. Regression Analysis

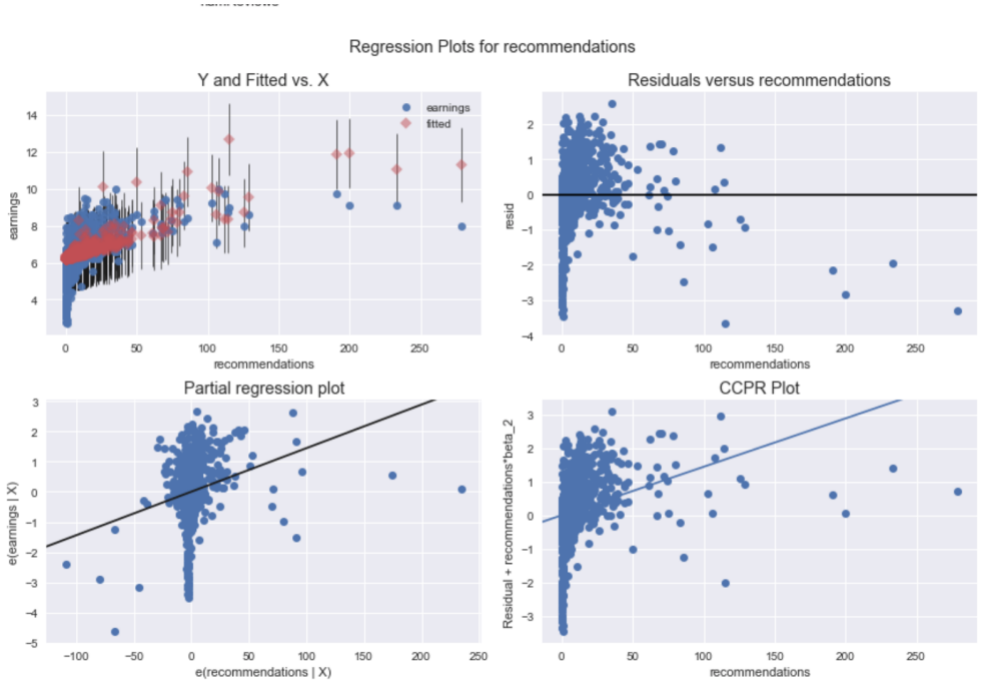
We Created a dictonary with users location(latitude and longitude), hourly rate, reviews and ratings. The code is attached as a jupyter notebook in the zip file

Sentiment Analysis:

1. We extracted six recent reviews of each freelancer and then performed sentiment analysis on it.
2. There are two files available with positive and negative words. We took all words from one review at a time and compared it with positive and negative words file.
3. This way we found that each review is either positive/negative/neutral.
4. After performing sentiment analysis on all extracted reviews of each freelancer, we decided freelancer's overall review.
5. If freelancer has equal number of positive and negative reviews, then overall review is neutral. If more number positive reviews then overall review is positive or else it is negative.

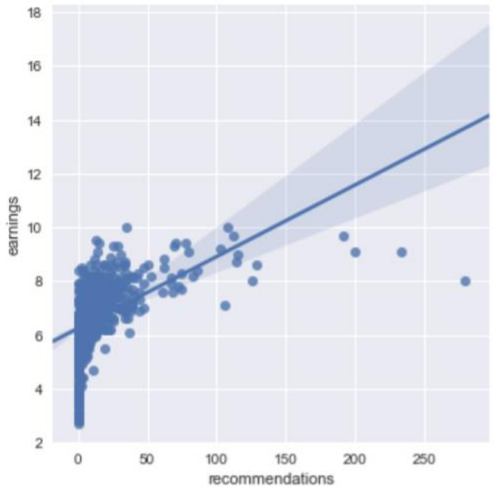
Regression Results:

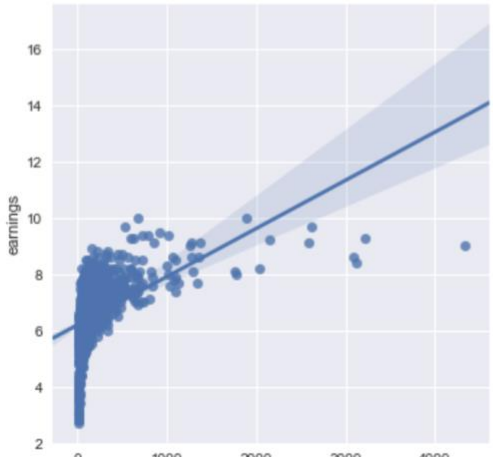
We applied regression , the following results are observed:

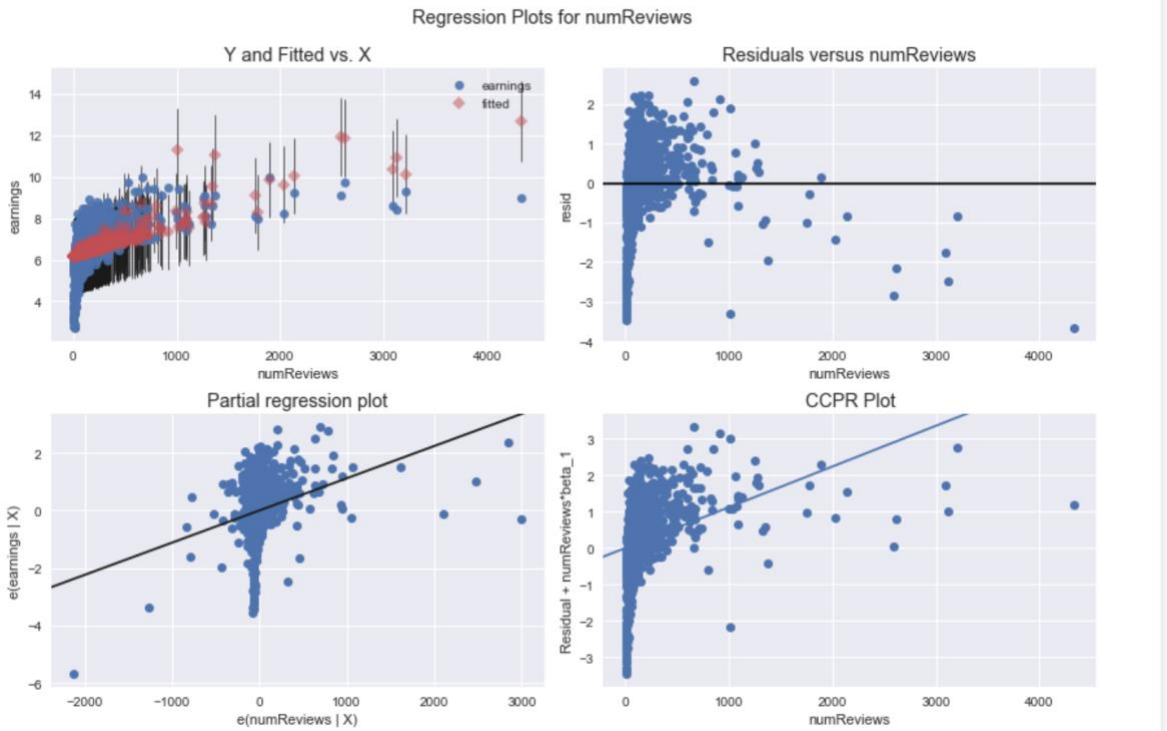


:

```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```







We can see from the model and graph that Earnings increases as number of reviews and recommendations increase.
However, the behavior is bit random when number of reviews, and recommendations are low.
For Higher values, the linearity can easily be established.

Future Scope:

We may need to apply some transformations (example, log) to make the model better.