Midterm Project

Fan Feng

**Personal Statement:**

I chose the dianpin datasets for my midterm project.

Dianpin is a very popular APP similar to YELP, but mainly rooted in China. Thousands of people daily use it to explore the gastronomic location of the city. In total, I have three open datasets containing the detailed information of the cities, the businesses and the reviews.

Here are the reasons that I chose the datasets:

My career vision is to become a data scientist who uses data to improve the quality of people's lives, including myself. As a food lover, nothing is more exciting than applying your professional skills to the things you love. Moreover, based on my experience, I firmly believe that these datasets are very interesting and worth exploring. A variety of cuisines and different regions allow us to group data in multiple ways and appropriately use multi-level models. Also, this project will strengthen my ability of handling complex real-world problems, story-telling with data and flexibly using multiple regression models.

In general, I can apply multiple techniques and models to play with the datasets and gain the answer of the questions that I really care about. This project gives me enough passion and will definitely benefit my future career.

**Question:**

What does the rating of a restaurant really mean?

What is the relationship between the ratings and the cuisines styles, regions and the number of comments?

**The data sources:**

The dataset can be freely used for academic purpose

<https://wise.cs.rutgers.edu/dataset/>

**Proposed Timeline of Work:**

EDA: 11/5 - 11/10

Data Processing: 11/10 - 11/15

Modeling and Validation: 11/15 – 11/20

Write up: 11/20 – 12/1