



K. N. Toosi University of Technology

**Faculty of Physics
Educational Group of
Atomic-Molecular and Astronomy**

Special Topics I Final Projects (Project 3 - Best Place)

Dr. Mohammad Hossein Zhoolideh

Academic Year 1403-1404

Overview

Currently, decisions about when and where to open new restaurants are largely based on the personal judgment and experience of development teams.

Setting up new restaurants requires a lot of investment and time. Statistics show that when the wrong place is chosen to open a restaurant, the restaurant closes within 18 months and suffers a lot of financial losses.

Finding a mathematical model to increase the effectiveness of investment in opening new restaurants can be very useful. In this project, we ask you to predict the annual sales of a restaurant in 100000 regional locations using demographic, real estate, and commercial data.

The description of each column in the dataset is as follows:

Id Restaurant id.

Open Date When does a restaurant open?

City The restaurant is located in the city. Take note that the names contain Unicode.

City Group Type of the city. Big cities, or Other.

Type The restaurant's type. (DT: Drive-Thru, MB: Mobile, FC: Food Court, IL: Inline)

P1, P2 - P37 These obscured data fall into three types. GIS-equipped third-party providers of demographic data are the sources of this information. These include the population in a particular area, the distribution of ages and genders, and development scales. Real estate data mostly pertain to the location's square meterage, front facade, and parking availability. The existence of sites of interest like banks, schools, and other QSR businesses is the primary component of commercial data.

Revenue Predictive analysis is focused on the revenue column, which shows the restaurant's (transformed) revenue for a specific year. Please be aware that the values are converted and do not correspond to actual currency amounts.

Note: The given data is raw. To answer this question, you must first preprocess the data using the Pandas package.

Important Points

Be sure to

- Leave appropriate comments for different parts of your code.
- Completely explain about the algorithm(s) you use to answer this question.
- Use **model selection**, **feature engineering** and **feature scaling** in your code.
- Measure your model performance using model evaluation metrics and interpret the obtained result(s).
- If you used a specific book or article in your project, mention it in your notebook.

A part of your score will be allocated to these items.

* You should write all the steps of your project in the **Jupyter notebook** and upload it as a file with the **.ipynb** extension on the vc site.