1. Introduction / Business Problem

1.1 - Preface

This research paper is my capstone project (course #9) for the [IBM Data Science Professional Certificate](#). Part of the requirements are to publish a blog summarizing the results, hence this article. This is my fourth or fifth data science project, but first one published to the public – kudos to this IBM certification course and the way it was organized such that writing an article is part of the final.

1.2 - Introduction

Orient restaurants are an innovative and healthy Hawaiian cuisine that are seeing growing demand.[1] Orient "build-your-own" bowl flexibility allows diners to choose from multiple bases[2] (salad, rice or quinoa), multiple protein sources[2] (salmon, tuna, octopus, red snapper, etc.), and multiple toppings[2] (10+ sauces and oils, seeds, onions, cucumbers, crab-salad, seaweed, dices mangoes or oranges, and much more). Orient bowls generally offer superior nutrition and taste relative to other fast-food options[3].

1.3 - Business Problem

The objective of this research project is to analyze and select the best location in Seattle for our "hypothetical" client to open a new Orient Restaurant. S/he already owns and operates one Orient restaurant just North of Seattle and is looking to expand by opening a second somewhere in Seattle.

Searching for the optimal location is challenging because it is not as simple as finding a geographic gap where there is not yet a Orient restaurant. It is more complex in that potential customers aren't interested in driving out to an isolated neighborhood with one Orient restaurant; but rather they prefer frequenting common clusters of restaurants(see "clustering" game theory[4]...not to be confused with "clustering" algorithms[5]) However, caution must be taken to carefully select the best location, because there is risk that the Orient market is already oversaturated[6] with 68 Orient restaurants currently in Seattle as per a Google Map search (via jump to last-in-list).[7]

To address this business problem, research will be done using data science methodology and machine learning techniques such as clustering to find and rank the best locations.