Applied Data Science Capstone Project

Route Recommendation System for Taxi Companies

Introduction

Taxis rely on being at the right place at the right time to maximise the number of trips they can perform per day. Unfortunately, it is impossible for taxi drivers to know for certain, when and where their next pick-up will be. Currently, ride-sharing apps like Uber and Grab provide taxi drivers with a matching service with customers such that they are able to drive directly to their customers when requested, rather than roaming the streets and hoping for a pick-up. This does increase their productivity.

However, this is highly subject to customer demand. When there is no ride request, taxi drivers will not know where to go. This raises the possibility that taxis are moving further away from their next likely pick-up point and increases travelling time to the pick-up point when a new request does come in. Therefore, it would raise productivity to be able to recommend a driving route for taxis, whether or not they are on a drop-off or an intended pick-up, to ensure that they are at the closest possible distance from the next probable pick-up.

In addition, taxi drivers may gravitate towards regions of high demand as they perceive that these locations provide the highest chance of a pick-up. While this is true, to maximise productivity, they need to increase trips per unit time. Remember that taxis are free-roaming and not tied to any particular route or central depot. They may be at any location at a given time, potentially far away from high-demand zones. As such, there is a need to evaluate whether it is worth it to travel 30 minutes for a 75% probability of a pick-up or 10 minutes for a 50% probability. Not many would choose the latter or even be aware of that option. Therefore, the recommended driving routes would have to flow through regions of increasing taxi demand, rather than making a beeline for the region of highest demand.

With an analysis of historic location data and behavioural patterns of a city’s residents, the neighbourhoods with highest human traffic (and therefore highest probability of a pick-up) can be determined real-time and even ahead of time. This would then be used to recommend an optimal route for drivers, based on their current and future locations, so as to minimise delay between pick-ups and maximise the probability of pick-ups per unit time.

Data

Historic location data from FourSquare will be used for my analysis. Even though not everyone checks-in on FourSquare or even uses the App, FourSquare users and their habit patterns can be considered a microcosm of the overall demographic of the city. As such, I will be analysing human traffic density and correlating that to demand for taxis (high human traffic = high probability of pick-ups). The following ‘Places API References’ will be used, in order of importance:

* Trending Venues. To observe human traffic in venues and neighbourhoods real-time.
* Similar Venues. Venues similar to those trending at the moment, are likely to have comparable human traffic due to similarities in habit patterns (e.g. smaller bars on a Friday night when a nightclub is the trending venue).
* Venue Categories. To derive patterns in human traffic by relating venue categories to date/time (e.g. restaurants during mealtimes, shopping malls during holiday season etc.) to predict upcoming trending venues.