OurBus
Assignment - I for Data Scientist

Himanshu Bansal
himanshubansal2005@gmail.com

Question 1

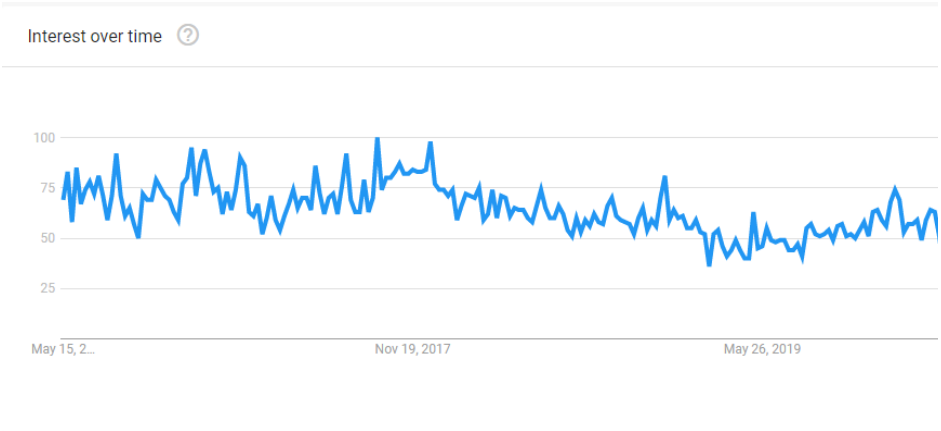
There are some days during the year when demand is exponentially higher than others, in the travel industry. How would you go about predicting this demand before the occurrence of such an event assuming that such an event is happening for the first time in the lifetime of the company?

Solution

To predict the forthcoming exponential demand, the information can be retrieved and analysed from the following sources:

- **Future date searches on the company's website**
    - Analyzing user's current and past ticket booking/surfing history to find out the future dates on which it is planning to book the tickets
- **Long Weekends / Weekends / Vacations / Festivals**
    - Demand is usually higher during these periods and the company can plan well in advance by looking at the calendar
- **Hotel bookings**
    - This implies only in the case when people stay in hotels while travelling. Company can look into the status of hotel bookings by scraping data from websites like OYO / MakeMyTrip to find out at which future date the availability of rooms is less

- **Average pre-booking period**
    - Usually there is a certain amount of time period when people start to plan their vacations/travel. The company can take into account the time period (let's suppose 10 days or 1 month) and subtract this time period from the data obtained from points 1,2,3 mentioned above.
    - Example: We came to know that people are searching for tickets on 20 May. We can subtract the average pre-booking period from this (10 days) from this date i.e. 20 May - 10 days i.e. 10 May.
- **Search trend of routes on google**
    - Google Trends can be leveraged to find out the trend of travel searches from the past years.
    - We'll be able to see the seasonality as well as the spikes of searches for travel bookings in past years.
    - For Example: Here is the search trend for Delhi to Jaipur route for the past 5 years excluding COVID-19 period.



"Delhi to Jaipur" search trend

We can clearly see that search trend is experiencing a spike in winter season i.e. November - January because of cold temperature and optimum weather for tourists to visit.

- **SEO and Digital advertisement data of competitor websites**
    - We can look at the advertisement trends that competitor websites have put in the past and also in current period to look for the upcoming demand
- **Similarity in route wrt. tourism and business**
    - If two routes are similar with respect to tourism or business purposes then data captured on one route can help forecast the demand on other route
    - For example the Delhi to Amritsar and Delhi to Jaipur routes are similar in tourism and business aspects. So forecasting demand on one route can help forecasting on a similar route as well.

Question 2

If you were given the responsibility to identify the 10 most profitable train routes in the Indian market, how would you go about approaching this problem and why?

Solution

To identify the most profitable train routes in the Indian market, we have to first look why is the route more profitable? If we succeed in finding out the factors which affect the profitability of the route then we can easily find out the most profitable routes. Certain data can be retrieved and analysed from the following sources:

- **Capacity utilization of a route**
    - Profit from a certain route is directly proportional to the utilisation of a train's capacity. We can also say that profit is inversely proportional to the vacant seats left on a particular route in near future.
    - Data of vacant seats can be scraped from **IRCTC,** which will allow us to know which route is being utilized the most and hence being more profitable.
- **Demand from Google search trends (https://trends.google.com/)**
    - We can focus on the highly searched routes on google (Example: "Delhi to Jaipur train")

    - This will allow us to compare routes which are more popular and busy than the rest of routes
- **Premium services routes**
    - Profitability of a route also depends upon the premium services available on a particular route.
    - For example Suvidha trains (formerly Premium trains) are fast Express rail service by Indian Railways which is operated in busiest routes. These trains have dynamic fare structure which means ticket price varies according to booking rate and seat availability. Priority of Premium trains are above Superfast and Mail/Express trains.
    - We can scrap this data from IRCTC and filter out which route is providing these high quality and reliable services
- **Connecting cities and towns**
    - Towns and cities connected by the route will play an important role in the profitability.
    - For example a route which is connecting metro cities will yield more profit than the route which is connecting a metro city and a rural or suburban area.
    - We can scrape this data from IRCTC and filter out the routes connecting metro cities and densely populated areas.
- **Routes which can be profitable but are currently not**
    - It has been noticed that people opt for alternative means of transport when the train route has unsuitable timings or poor quality services
    - For example: To reach Patiala from Delhi I opted for Bus rather than train because of multiple reasons including quality, speed and favourable timings
- **Routes with less number of stops provided that capacity is utilized fully at terminal stations**
    - Profitability of a route is inversely proportional to the number of stops in between the major towns or cities provided that the capacity is fully utilized at the terminal stations.
    - We can scrape this data from IRCTC and filter out routes having less number of stops in between and also the terminal stations have less number of vacant seats available for booking.