Datasets

Dataset 1: Zomato Restaurants dataset(zomato_modified). This dataset is made largely from an Indian perspective and we hope you'll enjoy the experience of analysing it.
The dataset is a list of different restaurants with a variety of self-explanatory features for each of them. It has been collected from the Zomato app and reflects some of the information they have.
Some questions you can ask here are:

1. Are the prices of a certain cuisine, say Italian, different in different countries? Why? Are their ratings different too?
2. Do places like Mumbai and Delhi have the maximum "fancy" restaurants, where fancy in some sense represents costly restaurants or may be restaurant with exotic cuisines?
3. Which cuisines are the most popular? Which are the most expensive? Within India, restaurants in which cities have the best food rating? Which cities are the most expensive in terms of dining out?
4. Is it true that restaurants with higher votes usually have a high rating? Do expensive restaurants have high ratings? What is your take on the correlations between these columns?
5. What is the average cost for two in different Indian cities? What is the average cost among restaurants that deliver via Zomato? Do most restaurants allow reservation through the Zomato app?

It is a challenge to you to go beyond these questions and demonstrate how you think about new datasets!
Bonus Points if:

a) You go beyond these questions and answer a few questions of your own.
b) If you are actually able to formulate a proper Machine Learning problem out of this and/or make an attempt to solve it.
c) You are able to formulate a Prescriptive Problem out of it and/or make an attempt to solve it

Dataset 2 : Hotel Booking Dataset(hotel_bookings). The dataset contains information about the stay of guests at two hotels – A City Hotel and a Resort Hotel situated in Portugal. Further Information about the features can be found at this link : https://www.sciencedirect.com/science/article/pii/S2352340918315191

Some questions which can be asked here are:

1. How many bookings does each hotel get every month? How many of these bookings are cancelled?

2. How do the number/ proportion of bookings vary with the market segment for each restaurant? Does the market segment have any correlation with the bookings getting cancelled?
3. Do the hotels cost the same? How does their pricing scheme vary with respect to the month and the market segments? Do you note any kind of special behaviour?
4. Has there been any change in the demographics (country of origin) visiting those hotels over the years?
5. How does the price charged by the hotel vary for the different kind of rooms offered for each hotel? Or even for different market segments?

Again, as was mentioned in the previous dataset, it is a challenge to you to beyond whatever we have asked and demonstrate how you think about new datasets.

Bonus Points if:

a) You go beyond whatever has been asked and answer a few questions of your own.
b) If you are able to formulate a Machine Learning problem out of this and/or make an attempt to solve it.