P1. Forecasting the Sales of a Supermarket During Festival Season

A supermarket has various departments, and it must stock up items that will be in demand in each of these departments. However, while stocking up, it must make sure that it does not have excessive stock, which it will not be able to ship out. Hence, you should be able to predict the impact of a festival season on the department-wise sales of a supermarket.

First, you can use a dataset from Kaggle, and for executing the project, you will need to choose a given holiday, let's say Christmas. Then, you will have to check if, during the time of Christmas, the store marks the highest numbers in sales and which departments need to stock up more items to meet the rising demand.

P2. Sales Conversion Optimization

A company does a lot of marketing to get sales, and various kinds of campaigns are initiated to market products. Campaigns such as email blasting and social media marketing are among the most popular ways of marketing a product.

The aim of this project is to understand what the most effective ways are in terms of ROI (return on investment) and which campaign generates more leads and then suggest the ways of going about this marketing campaign in the most optimized manner based on a provided budget.

For this Business analytics project, you can use the following dataset to get information on a company's marketing campaign data.

P3. Employee Attrition and Performance

A company wants to understand what factors lead to employee attrition (i.e., it is trying to know when and why an employee decides to leave the company). By understanding these factors the company wants to change its business environment accordingly so that it can hold on to its best employees.

In this project, you will need to evaluate each factor and its relationship with attrition, for example, the distance from home to office, the job role impact on attrition, etc. For the dataset, you can click here and carry on with your evaluations.

P4. Predicting Sales in Tourism for the Next 4 Years

Tourism is one of the fastest-growing industries in the world. With the introduction of hashtags like '#wanderlust', there has been an increasing amount of interest among people of different demographics to explore new places. However, this industry has very fluctuating numbers in terms of sales, and different places have different feelings according to the time of the year.

Hence, tourism forecasting has become an increasingly important task in planning, improving, and managing the industry. There is a lot of information and insights that are hidden in the data retrieved from the tourism industry. You can use techniques like data clustering to understand when and where tourists prefer to go, what they like at each location, the mode of transportation of tourists while travelling between spots, etc.

Using insights like the above, you need to forecast the sales for the upcoming 4 years. You can use this dataset for your evaluations and then compare them with the actual data.

P5. Predicting the Success of an Upcoming Movie

The entertainment industry has been growing in every scope. Be it Netflix, Amazon, or Hotstar, there is a lot of content out there. Now, the challenge these streaming services face is what to buy, in the sense that which content will get them more viewers and also satisfy the existing customer base.

For this project, you need to predict the success of an upcoming movie so that whether or not a company should go for buying it based on ROI. To do this, you need to come up with a model and use the historical data of each element involved, such as the actors, the director of the movie, the production company, the genre of the movie, etc.

The main idea of doing this Business Analytics project is to predict the market for the upcoming media content based on some preset parameters as this is one of the most unpredictable industries: Big stars might not always shine, while the newcomers might actually do a great job! You will need to keep all of that in mind.

Data Collection: The first step was to collect data. There was no pre-existing dataset with the movies and features we wanted, so we collected data from multiple sources and merged it. Our main sources of data are IMDB, Rotten Tomatoes, and Wikipedia. The mechanism through which each of these sites makes their data available varies everything from an API to messy space-delimited text files.

From IMDB, we obtained for each movie: movie title, IMDB rating, plot description, budget, box office gross, and opening weekend gross. From Rotten Tomatoes, we obtained critic score, audience score, runtime, MPAA rating, studio, theater release date, DVD release date, list of genres, abridged list of cast, and abridged list of directors. From Wikipedia, we obtained the number of Academy Awards that actors and directors in each movie had won prior to that movie, and also the number of Best Picture films that actors and directors in each movie had been involved in, also prior to that movie. We collected the same data from the Golden Globes. In order to stay true to our goal of only considering factors known before a movie's release, we considered only awards that had been received prior.

Using a data processing tool (Hadoop/Spark, Python, JavaScript) to parse, and clean up what we retrieved, we had our data – albeit unmerged. We removed rows from each table that did not have a complete tuple of information (in particular, Rotten Tomatoes often had important attributes missing about less popular movies, like runtime or studio). To merge the files, we keyed on a combination of two fields: normalized movie title and release year. To normalize movie titles, we uppercased all titles and stripped all non-alphanumeric characters. Finally, we loaded each table into an SQLite database and joined all three tables, leaving us with a data set of movie information with 6,590 movies. Our data made sense: the highest rated movies across IMDB and Rotten Tomatoes include "The Dark Knight" and "The Godfather" and the lowest rated movies include "Justin Bieber: Never Say Never" and "From Justin to Kelly". Alas, maybe the best classifier predicts that if a movie title includes the word "Justin", it is destined for failure.

Our data set represents many features as bit vectors. Some features have been parsed into other features, such as splitting the release date into bins by month and bins of five years. We also added a feature for whether a movie came out on a popular weekend (like the Fourth of July or Christmas), and included a feature for whether a movie was a sequel, third installment, or later in a series. Our final input feature vector contained 176 features: 6 pertaining to bit vectors for MPAA rating, 1 for movie runtime, 22 pertaining to bit vectors for movie studio, 12 for release month, 4 for popular weekends, 8 for year bins, 1 for budget, 6 for awards (one for each of the following for both Academy and Globe: sum of all award winners among the cast, director, and those involved in a Best Picture), 22 pertaining to bit vectors for genre, 90 pertaining to bit vectors for the most popular words in a movie's plot description (after removing stop words and words that seemed unrelated to movie content), and 4 bit vectors for the movie's place in a series.

P6. Customer Segmentation

An e-commerce company has a variety of customers. Every customer has a different set of tastes and interests and may belong to different financial levels. Therefore, it is a heavy challenge for the marketing and strategy team to decide what products it should be promoting or what kind of campaigns will lead to the most lucrative results.

Spending score is one such metric through which you can segregate customers. Spending score is not just determined by the financial situation, but it is also based on other factors such as customer behavior, the kind of products a customer buys, etc.

In this project, your marketing team basically wants you to identify the customers who will most easily converge and buy products. In doing so, you must show the different segments in percentage and also predict the kind of products and marketing campaigns that will be the most successful with your customers.

In this one of the best projects on Business Analytics, you can use this e-commerce dataset and mall customer dataset.