

Module wise Case Studies List

I. Data Exploration (R, Python Programming and Statistics)



Start-up Success Rate and Investment Analysis

The data contains information about the start up as well as their founders and co-founders. Data analysis can be carried out on the data to help investors identify important factors that contribute towards the success of a start-up to aid them in their investment strategies.



Insurance Premium Analysis

The health insurance companies remain profitable when the premium paid by the people is more than their medical charges. The data need to be analyzed to identify factors that lead to less medical charges.



Consumer Goods Sales Forecast

Sales analytics is the process used to identify, model, understand and predict sales trends and sales results while helping in the understanding of these trends and finding improvement points.



Automobile Price Analysis

Automobile data involved analyzing factors that affect the price of the automobile. The dataset consists of information about the body style, make, fuel type etc.



App Market Analysis

The Google Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market.



Analysing Flight Operations

This dataset contains all flights departing from Houston airports IAH (George Bush Intercontinental) and HOU (Houston Hobby) in 2011. Various data manipulation operations can be performed to get insights about the flights schedule, status, route etc.



Movie Recommendation

The datasets describe ratings and free-text tagging activities from MovieLens, a movie recommendation service. Analysis is done to answer different questions about the recommendation of the movies.



Real Estate Price Prediction

Ames housing dataset consists of 79 explanatory variables describing (almost) every aspect of residential homes in Ames. The goal is to predict the price of each home.

II. Machine Learning



California Housing Price

The data contains information from the 1990 California Census. The data pertains to the houses found in a given California district and used to predict the housing prices for a district.



Health Care Quality Assessment

The dataset consists of the claims data for the insured patients and the goal is to predict the quality of healthcare received by the patient from the medical provider.



Baseball Players

This dataset was originally taken from the StatLib library which is maintained at Carnegie Mellon University. The data is analyzed to predict the salary of baseball players playing in major leagues.



Heart Disease

The dataset consists of 14 variables and is to identify trends in heart data to predict certain cardiovascular events that can lead to chronic heart disease.



Movies Recommendation

The case study aims to build a movie recommendation system based on genres. The data used is gathered from MovieLens.



Automobile Sales

The case study predicts monthly sales of the Hyundai Elantra in the United States.



Framingham Heart Study

The Framingham Heart Study is a long-term, ongoing cardiovascular cohort study of residents of the city of Framingham, Massachusetts. The data is used to predict and prevent heart disease.



Mall Customer Segmentation

The Mall Customers data set contains information of 200 customers and their annual income and spending pattern. The data set can be used to carry out customer segmentation



Credit Risk Modelling

Banks and other credit providers need to understand the likelihood of loans being repaid. The dataset consists of the loan data and the output is to identify whether a person will be able to pay/default the loan amount

III. Deep Learning



MNIST Digit Recognition

MNIST digit is a dataset of 60,000 small square 28×28 pixel grayscale images of handwritten single digits between 0 and 9. The digit are classified into the 10 classes using Artificial Neural Network



Fashion MNIST

Fashion-MNIST is a dataset of Zalando's article images consisting of a training set of 60,000 examples. The dataset consist of the images of various fashion articles like skirt, t-shirt, socks etc. that are classified into their respective classes.



Movie Reviews Sentiment Analysis

This is a dataset of 25,000 movies reviews from IMDB, labeled by sentiment (positive/negative)



Image Recognition on CIFAR-10

The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. The dataset is used as a benchmark for computer vision algorithms.



Image Classification on Flowers Dataset

The dataset consists of flower pictures that are divided into five classes: daisy, tulip, rose, sunflower and dandelion. CNN is used to classify the flower images in respective class.



Cats and Dogs classification

The dataset consists of 50000 images of cats and dogs. The task is to classify cats and dogs with the help of Convolutional Neural Networks.



Cards Recognition and Detection

The dataset consists of the images of nine, ten, jack, queen, king and ace cards from a deck. The goal is to identify the card and detect the card in the image.