

Final Project 1

Data Science

Project 1 - College Scorecard

Problem Statement

Scorecard dataset provides information around college cost and outcomes by year so that the students and parents can weigh and compare different options.

Using dataset what could be the average cost & debt for students after the completion of the school. Is there an relationship between the cost incurred and how well the student perform?



Project 1 - College Scorecard contd...

Data

This data set has information about the institute, Location, Type of program, Acceptance rate, Cost, Aid, repayment and Earnings data.

Most recent dataset was extracted from the gov website. This includes data from 1996 to 2013.

(1) <https://catalog.data.gov/dataset/college-scorecard> (2) <https://collegescorecard.ed.gov/data/documentation/>

Hypotheses

We are assuming that all the aid was provided through the financial aid program and that there are no other external borrowings



Project 2 - Online News Popularity

Problem Statement

This dataset summarizes a heterogeneous set of features about articles published by Mashable in a period of two years. The goal is to predict the number of shares in social networks (popularity). This would help both the writer and publisher make a profitable choice

Data

The dataset has attributes like url,tokens, unique token, images, channel, published day, polarity of positive words, title polarity, number of shares

Dataset Provided by - <https://archive.ics.uci.edu/ml/datasets/Online+News+Popularity>



Project 2 - Online News Popularity contd...

Hypotheses

We are assuming that number of shares would help us predict the popularity



Project 3 - Traffic Violations

Problem Statement

The traffic violations has information about all the electronic traffic violations issued along with additional attributes like location, agency reported, accidents, time, type of etc.

Using the data can we predict if the accidents occur based on certain attributes. What factors can we change to avoid this?

Data

This dataset has information on when and where the violation happened along with attributes like Gender, Accident, Alcohol consumption, vehicle type, Race, Property Damage

Dataset provided by - <https://catalog.data.gov/dataset/traffic-violations-56dda>

Project 3 - Traffic Violations contd...

Hypotheses

Since the tickets/violations were captured electronically, we assume that there wasn't any bias

