Final Project Ideas

Elevator Pitch

Problem 1: Why We Quit Our Jobs

Employee attrition affects every workforce around the world. Companies losing their top talent can cost them a significant amount of time and capital to replace. Figuring out why employees leave their jobs, and being able to predict which employees are at risk is valuable to any Human Resources team.

Potential Data

Two data sets on Kaggle contain fictitious HR data to use

- IBM’s data contains education, commute distance, job satisfaction, job performance, and other unique, potential predictors
- People HR Analytics Repository contains 10 years of ‘data’, termination reason, length of service, department, and store number. This allows for analysis under a single business.

Hypothesis

Given the data I will find that attrition is most affected by commute distance, work life balance, and work environment than salary, standard hours, and overtime.

Problem 2: Which games are trending?

Video games are a $99.6 billion dollar industry with steep monetary and human development costs. With the price of each unit not changing since 2006 and the cost of development skyrocketing yearly sales trends are incredibly valuable. Figuring out which kinds of games will sell millions of copies is paramount for publishers and developers to stay in business.

Potential Data

- Sales data scraped from VGChartz compiled with corresponding data from Metacritic
- More than 16,000 records of games which sold 100,000 units
- More than 6,900 records with ratings and sales data
- Data obtained via Kaggle

Hypothesis

Total sales will be best predicted by which platforms the game is on, and which genre it belongs to.

Problem 3: QA Consistency

Production and manufacturing has the ability to bring value-priced but high-quality products to the masses. However, losses along the process or decreases in end quality hamper profit and reputation of companies. Being able to find where in the manufacturing process these events occur will minimize profit loss both internally and loss once low quality items hit the market.

Potential Data

Bosch has compiled a massive amount of data where it notes the quality of products along each part of production.

- Thousands of measurements are made at each step
- Large enough that the data is broken into 3 datasets
- Each part has a unique ID to follow through the process

Hypothesis

The problem areas in the production process will be found and the amount of lost parts will be reduced by 10%.