Information Seeking

Dataset 1: Stack Overflow Survey Response

Stack overflow has helped developers solve problems, grow their careers and learn new skills. The chosen dataset is a survey response of over 50000 software developers collected by Stack Overflow to come up with additional insights. The data has comprehensive details ranging from Occupation, Current employer details, hobbies, how often do people use stack overflow, expectations from the employer and reason for using stack overflow. This dataset can help in predicting job seekers, high potential employees for recruiters. It can also help in predicting new employee benefit programs for IT organizations based on the expectations of the employees in the dataset.

Potential data users and decision-makers:

- Job recruiters
- HR Team in organizations

Questions:

- Predicting job seekers using stack overflow.
- Relationship between current work profile and technology the employee wants to work on in the current year?
- Relationship between years of programming experience and use of stack overflow?

Information Seeking

Dataset 2: Consumer Complaint Database

Consumer Financial Protection Bureau. (2013, November 06). *Consumer Complaint Database*. Retrieved September 13, 2016, from Search for a Dataset - Data.gov, https://catalog.data.gov/dataset/consumer-complaint-database

License: License not required

Terms of use: https://www.kaggle.com/terms

The dataset published by the Consumer Financial Protection Bureau consists of complaints received about financial products and services between 2011 and 2016. The data contains details of the complaints like the type of product, category of the issue, company, location and date. Customer complaints is a form of feedback which can be used to make interesting predictions like whether measures were taken to target specific issues by companies to ensure that the count reduced over a period of time.

Potential data users and decision-makers:

- Commercial Banks
- Investment Banks
- Insurance Companies
- Brokerages

Questions:

- Relationship of number of days taken to send the data to the company and the type of issue in a product?
- Identify the frequency of specific issues in products annually to check whether the count was lesser in the consecutive years?
- Relationship between complaint from customers about issues in a product and States in US to identify a pattern.

Information Seeking

Dataset 3: Payment default of credit card clients

Yeh, I.-C. (2016, January 26). *UCI machine learning repository: Default of credit card client's data set*. Retrieved September 13, 2016, from UCI Machine Learning Repository: Data Sets, https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

License: License not required

> The link gives details regarding terms of use for data on the website:
> http://archive.ics.uci.edu/ml/about.html

The number of transactions executed using credit cards are way higher than that of transactions done in cash. According to survey details on http://www.creditcards.com/ given below, the statistics show that the preferred method of online shopping payment is credit cards. The data set consists of credit card payment history details for individuals in Taiwan for the year 2005. The data contains details regarding the limit balance, personal details, bill amount and amount paid for a period of 5 consecutive months. The increasing use of credit cards and predicting the risk of customer credit default by analyzing the data makes the data set interesting.

Online shopping preferred payment Method		
Payment type	**2013**	**2014**
Credit card	41%	48%
Debit card	22%	30%
PayPal	22%	12%

Potential data users and decision-makers:

- Banks
- Credit card clients for predicting their own default payment

Questions:

- Relationship between human factors (Sex, Education, Marriage and Age) and payment default?
- Does higher limit balance on a credit card predict higher frequency of payment default?
- Predict the risk of customer credit default?