**Credit card default**

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 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% |

Questions:

* Relationship between human factors (Sex, Education, Marriage, 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?

**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

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.

Questions:

* Relationship between complaint from customers about issues in a product and States in US to identify a pattern.
* 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?