IDS 570
Statistics for Management

Homework 4: Project Proposal

Team Members:
Ashutosh Dwivedi (674351427)
Mohak Behal (672119300)
Mohammed Rehan (668333717)
Pavan Palve (670397922)
Srija Gupta (656331239)

1. Identify a data set relevant for the statistical analyses that your team expects to use for the Final Project.
a: How have you obtained the data?
Answer: Dataset is a sample of data maintained by one of the auto-insurance providers named Indian Money, Bangalore, India. One of our friend is working for this organization and we have received the data from her.

b: How were the data collected? For example, an employee entered a code during a call with a client, based on the attached Customer Follow-Up Form. Survey data downloaded from data.gov (with a link to the data description).
Answer: Data is collected by the organization during claims processing and reporting of claim data at the end of the year.

c: What are the main characteristics of the data?
Answer: The dataset has 15 variables and 7702 observations.

Variable	Datatype	Description
Policy Number	Numeric	Unique Policy Identifier
Year	Factor	Year of manufacture
IDV	Numeric	Insured Declared Value of Car
City	Factor	City of registration of vehicle
State	Factor	State of registration
Cubic Capacity	Numeric	Capacity of engine
Mfr-Model	Factor	Manufacture and model of car
Premium	Numeric	Total Premium paid for the policy at the beginning of the term
Type	Factor	Source of lead
Gender	Factor	Male or Female
Channel	Factor	Lead generation channel
Age Group	Factor	Age Group of applicants

Payment Frequency	Factor	Annual payment or monthly instalments
ClaimsInd	Factor	Claims taken (0 - Not taken|1 - Taken)
Claim Amount	Numeric	Claims Amount

2 Elaborate the following:
a: What is your research question?
Answer:

"What factors affect profitability of Indian Money for NCR region in the year 2011?"

b: What are the dependent, independent and control variables?
Answer:

Dependent Variable: Profit.
Profit will be a derived from column IDV, Premium and ClaimAmt variables
Independent variables: Age Group, Gender
Control Variable: State, Year

c: What are your hypotheses?
Answer:
"The profitability of Indian Money for NCR Region in year 2011 is higher for female drivers having age more than 40 years"

D: What challenges do you anticipate experiencing during this project?
Answer: The challenges which can come in the process of modeling our data would be that profitability is affected by multiple factors but we need to screen out those factors which are most crucial in determining the profitability of our Insurance company.