

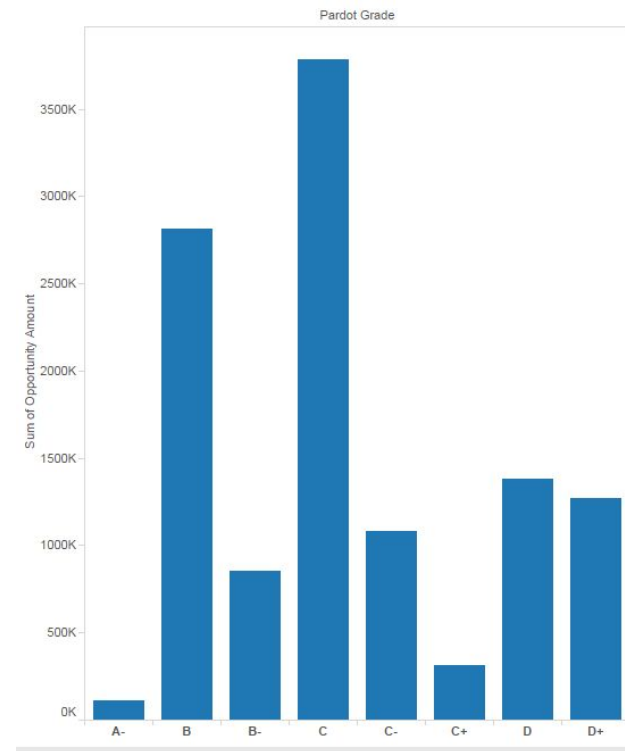
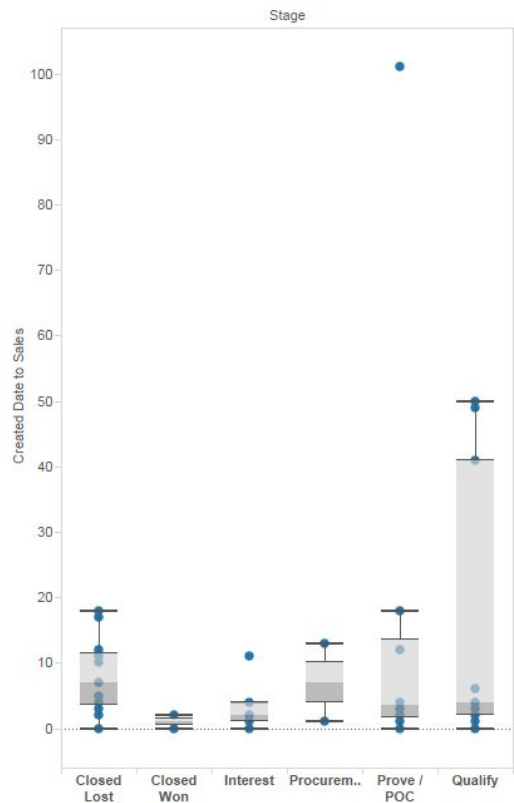


Data Understanding, Data Preparation, Modeling Plan

Data Description



- Lead 32 variables
 - Title
 - Pardot Scores/Grade
 - Lead Source
 - Lead Status
 - Company
 - Stage
 - Conversion Dates
- Campaign 16 variables
 - Member Status
 - Campaign Name
 - Response Type
 - Response Dates



Feature Creation

- Job Titles (Dummy Variables):
 - Business vs Technical
 - Board - CSuites - Consultant

Sum	29	17403	29112	4612	1362	2231	130	45374	15652
Title	board	c-suite	director	head	consultant	business	student	technical	others

- Company Name
- # of Leads
- Level of Engagement (Weight)

Missing Data

- Missing Data is dealt with in one of four ways:
 - Meaningful missing values: Member First Responded Date, Opportunity ID, Various Date Variables
 - Assuming the mode: Country
 - Assuming the mean: Pardot Grade/Pradot Score
 - Unimportant: Email Domain, Website
- Data Verification

Different Datasets

- Three different datasets: Leads, Opportunities and Campaign
- Matching on Company Name
- CSV formatted

Data Structuring

- Inner joins on all three datasets
- Using R to construct data frame
- k-fold cross validation to have train and test data sets
- Classification Problem

pprenda

Questions?