Brain Storming 15/2/17

- Hypothesized that Motorcycle and Car have very different nature of data and should be split for analysis
- As our dataset is very large, about 3GB, we have problems running analysis efficiently. Thus for expediency and also for our findings to be relevant, we limited our dataset to 2012-2015. This made the data size to be about 300MB.
- Combined GWP, DISC,COMMWritten_PREM
- Discussed on focusing on temporal pattern and profitability over time
- Discussed on trying to visualize high dimensional data using tools like parallel plots, ternary charts etc.

To Follow up for consultation:

- Present exploratory data analysis and garner feedback
- Talk about alternative visualization methods
- To discuss about prediction modelling.

Steps taken to Merge claims dataset:

1. Filtered Orig_InceptionDate by 2012-01-01 – Present
2. Combined GWP, DISC,COMMWritten_PREM
3. Removed Unused columns
4. Merged motor_policy30-2012start with motor_claim7_combined-2012 start . Joining columns **Policy** and **Risk_NO.**
5. Used Table-> Groupings: Risk_NO, Policy,Claim_NO Summary statistics(Sum)
6. Deleted **Transaction_Desc, TransCode_Desc,Payer_Name**
7. Deleted **Anniversary** from **motor_final**
8. Joined **Motor_final** With **Motor_final_summed** with matching conditions **Policy, Risk_NO**