**Projects’ Summary**

Cyber Survey Analysis

* Worked with Kara Owens and Ashley Dillon to try to find another input in determining what product divisions are more at risk to cyber exposure
  + Add analysis to which kinds of products we would like to mark as “affirmative cyber”
* Followed quantitative questions’ scoring and weighting methods to find total risk scores for each survey response from a PL
  + Bucketed each response into green, yellow, or red risk
  + Scores ranged from 0-57
    - 0-20
    - 20-40
    - 40-60
* Aggregated by average risk score with respect to the product divisions
  + Markel Assurance Commercial – 31.86
  + Markel Specialty Commercial – 29.06
  + Markel Global Re – 24.20
  + Markel Assurance Risk Managed – 22.0
  + Markel Specialty Other – 9.00

Microsoft Azure Text Analytics Service (IP Sprint)

* Went through Azure fundamentals tutorial and developed a basic understanding of the cloud
* Was tasked with finding risk pockets in a dataset of binding E&S homeowners policies that fell outside standard risk guidelines at the same time of IP sprint
  + Required text analysis + mining
* Performed sentiment analysis and extracted key phrases from text
* Service is not best equipped for insurance text
  + Azure has its own pre-defined parameters for finding sentiment or key phrases that are more geared towards things like twitter
  + Lot of instances in which text had high sentiment but still declined
* Could be a powerful tool for product feedback in which users can report a bug or inconvenience they experience with the product (underwriter workbook, claims dashboard, etc.)
* Service can automate process of finding sentiment and ranking in order of lowest sentiment to generate priority

Binding E&S Personal Property Analysis and Modeling

* Tasked to find appetite guidance that can be provided to producers to reduce the number of referrals and create more time for the underwriters
* Data challenges
  + All vital info is in two free text field columns (*Risk Details* and *Reason for Decision*)
  + Approval/decline classification is imbalanced (Approval ~85%, Decline ~15%)
* Performed text analysis on same binding property dataset with Python tools/packages
* Generated wordclouds for the *Risk Details* column for approvals and declines to get a feel for trendy words in the text
* Performed a bigram analysis (2 word phrases that showed up frequently)
* Used regular expressions to text mine the year that these houses were built and plotted the distribution for approvals and declines
* Also mined the response time and plotted the distribution for approvals and declines
* Created a document term matrix or “bag of words” of the *Risk Details* text which takes every single word in the column and makes it a feature
  + Applied a naïve bayes model on this matrix after resampling to come out with a weighted f1 score of about 0.75
* Created an attribute column for each key risk (brush status, loss history, year built, response time, and alarm system status)
  + Applied a decision tree after resampling to come out with a weighted f1 score of about 0.60
* Created visualization of the algorithm’s decision path

One Time Underwriter Workbook Analysis

* Wrote a query with the help of Nick Blankenship to pull the rolling 9 month information on policies from *Renewal\_Optimization.XRO\_Policies* to perform a one time workbook analysis
  + Gauge the impact of the workbook on business value by looking at metrics like total policy counts, time savings, and GWP acquired per minute
* Plotted total policy counts by product line
* Plotted average GWP by product line
* Used time savings data acquired by Nick Blankenship and Miranda Elliott to make estimates on total time saved and GWP acquired per minute
* Hundreds of hours saved and thousands of GWP dollars acquired