Microsoft Azure’s Text Analysis Service

Azure offers a useful text analysis service that could be useful for Markel when Markel has to deal with descriptions in their data. More specifically, in a personal lines dataset indicative of binding decisions, there are two data features that contain vital information to the binding decision – **risk details** and **reason for binding**. Data collection methods are never perfect and Markel may have to use some sort of text analysis if the situation demands it.

I was able to witness the abilities of Azure’s text analysis service by using it on such a dataset. The setup for this involved the creation of a resource group and cognitive service in the Azure Portal. The catch with this is that Azure’s services are usually not free. Although I was able to choose a free pricing tier when I purchased the text analysis service, I ran out of the allocated number of API calls (5000) for 30 days.

After securing the service I was able to use the Python REST API to call the text analytics cognitive service. You can find more info about this [here](https://docs.microsoft.com/en-us/azure/cognitive-services/Text-Analytics/quickstarts/python?toc=https%3A%2F%2Fdocs.microsoft.com%2Fen-us%2Fpython%2Fazure%2FTOC.json&bc=https%3A%2F%2Fdocs.microsoft.com%2Fen-us%2Fazure%2Fbread%2Ftoc.json). Although this method of text analysis requires an additional step of converting your raw data (excel) into a JSON format to be read by Azure’s cognitive service, I was able to find the **sentiment scores** and **key phrases** of the risk details and binding reasons in about 20-25 lines of code. Azure will automatically do these things for you after you simply call their methods on the data. Performing text analysis with a different method may not be as simple. **I would say that Microsoft’s library may not be as vast as other text analysis libraries but is still powerful and easier to use.**

I verified a few of the sentiment scores by individually looking at the text of policies with high sentiment scores and most of these policies tended to renew. One example is *“Was a member of band, most hits in 70s/80s. Nothing recent no concerns. Confirmed with LT OK to offer w/ liability.”* This policy had a sentiment score of about 98%, meaning likely to be approved within our context. I was also able to successfully extract important key phrases from all of the policies. However, given more calls I would be able to finish the analysis and come up with a more conclusive understanding of the sentiment scores and key phrases.