**CAS Applied Data Science Module 5 - Peer Consulting**

I got together with my colleague Casimir von Arx on Thursday, 30.05.2019 to consult him on his project. He programed a QlikView tool that downloads election results from the Bernese Grand Council and he tries to get as close to real-time forecast as possible during election day. His wish is to have the tool ready by 20.10.2019, which is national election day.

The model should take previous elections into account as there is always a size bias present. Small and conservative circles are counted out much earlier than the largest circle, which is Bern itself and rather liberal. So just predicting the elections with the current data would result in an estimate that would be significantly off each time.

**5 Good Things**

First, we will list five of the good things about Casimir’s project.

1. **Domain Expertise**  
   Casimir has great domain knowledge on many levels. This is not only apparent when one views his model, he can also answer to almost all inquiries and explain how things work. While this may be taken as granted, it is of utmost importance for a consultant to being able to draw on this knowledge.
2. **Naming and Clarity of Code**Casimir’s model is very intuitive to understand and to follow because the naming of the constants follows a clear and easy to understand logic. It is often the case that the naming logic is built up along the way and that people don’t bother to go back and align everything in retrospection. Not so here. It is another thing that might be taken as granted and it might be viewed as something of minor importance. But there are many examples of projects that had do be completely redone because the naming was obtuse and the code was not commented or not logically built up.
3. **Clarity of Vision**  
   Casimir has a clear idea of what his tool is supposed to do and where he wants to take it. He can clearly state the problems he ran into and he can also describe what he has already tried to solve or circumvent these problems. There is a good foundation to build consulting on.
4. **Expandability of Code**One aspect that is great in QlikView is that variables can be input at various places. Casimir did that to a high degree and very view things are hard coded, which makes the software easy to expand with new concepts and new code.
5. **Working Code**  
   Casimir’s tool is working for the test case he designed. We can fully focus on where to take the design from here and draw on many working parts. This is very beneficial in consulting as we do not need to go back and fix existing code.

**5 Things That Could Be Improved**

What Casimir has built so far is functioning very well already and there are only minor things that could be improved. However, as he aims to expand the range of his tool, I can make a few helpful suggestions.

1. **Automatic Fetching**  
   Currently, fetching the data is triggered manually. Since the workload on national election day will be much higher, it is advisable to automate fetching the data and having as close to online data analysis and prediction as possible. This includes changing some of the parameters of the download. As of now, download of the data overwrites the previous data, so there is always a full load of data. This works because we are only dealing with one canton at the moment. But when we expand to 26 cantons, only loading a delta is preferable.
2. **Increase Speed**  
   For a few reasons, the tool is rather slow when downloading the data. One aspect is that always a full load is triggered, which I have already addressed above. Another one might be that the code could be optimized in this regard. One suggestion is that instead of accessing data directly online, it might be advantageous to download the data, import it into QlikView and then run analysis.
3. **Confidence Intervals**Right now, the tool gives estimates of certain outcomes. What is missing are confidence intervals to give a sense of how good these estimates are and to give upper and lower boundaries.
4. **Neural Network**This case is perfect to implement a neural network and past data is available. One challenge is that we may not have enough data to train the network efficiently. However, this can be easily remedied because we can take the past data, split it up and do a cross-validation, trying to predict the missing piece from the pieces that are known. Another strategy might be to predict Bern as this is a very important function of the tool.
5. **Predict People**Currently, the tool tries to predict party success. An interesting way to go might be to try to predict single candidates’ success and then build the party success from this. However, this might be much harder as you cannot draw on past data for all candidates.