Hack Exercises

Phillip, Ramin, Anne-Marie

3 Sep 2019

A14 & USIT data

On accessing a PostGreSQL database from AWS RDB in PowerBI

- Install the PostgreSQL ODBC driver Install
- From the ODBC Data Source Administrator (type 'ODBC' into Cortana etc.) select [New]
- Select PostgreSQL Unicode(x64) from the list of options
- In the setup screen, select: – Data Source = PostgreSQL35W – Database = mcfc01 – Server = mcfc01.c4hbjfyzingo.eu-west-2.rds.amazonaws.com – Port = 5432 – User Name = mcfcuser01 – Password = projecthack01
- Add the Data source – hit [OK]
- In Power BI desktop, select Get Data | Other | ODBC
- Select the Data Source Name you just created above, e.g. PostgreSQL35W
- Re-enter user name & password when prompted
- Hit [Connect]
- In the Navigator you should see under ODBC the database mcfc01 with some public tables: A14, R2,etc.

There are a number of tables:

- agency_alpha
- bureau_alpha
- etc..

Each of these is the corresponding spreadsheet from the download, but with all the years concatenated (and a 'year' column added). Data has been converted to numeric where the process could be made automatic, but much of it still remains as character.

There are two views of interest:

- project2activity
- project2business_case

These are the simple join of the project spreadsheets with the activity spreadsheets and the business case spreadsheets respectively. I am looking through these to try to categorize projects using ML (e.g. Support Vector Machines) but have no results to speak of yet.

Let me know if you go through the data and find things that could be cleaned up or improved.

The A14 web app

- http://35.177.16.6:5000/mcfc
- https://github.com/possibly-sam/mcfc02

(I took the server down briefly – let me know if you want me to put it back up.)

Things that could be extended in this include:

Routine visualization extensions

- View how the risk changes over time, by reporting month
- Add the 'black swan lever'
- Can we create a mobile interface to allow entry of risks?
- Add comparisons against historical data or data from other comparible projects

Visualizing USIT data

- Add charts that display USIT data.
  - Cost Variance (absolute or % – both) for different categories of projects as a function of total cost, agency, NLP keywords
  - Get distribution of USIT data to determine if the risk distribution is good

Things that would require more information

- Can we get risk registers from the USIT data? Or Historical data from other Highways England projects?
- Is there a connection with the issue log?

Further advanced exercises

- What, in the USIT determines the cost variance ? Create a model to predict this.
- How would this apply to the A14?

- What makes two projects similar? Size, cost, duration, riskiness?
- What are predictors of an underestimated project
- Re-engage with Highways England to get a feel for how useful these ideas may be.
  - Or even just cleaning up the data