**Risk and Regression Data Test – PwC**

**Background –**

The flow of monthly payments is insightful for organizations in terms of revenue metrics and resource optimization. Many superannuation organizations are interested in knowing monthly cash flow. Here the goal is to design a report that gives aggregate monthly statistics on employer accounts and their superannuation payments, per month and per “employer tier”.

**Datasets –**

Two files were provided:

* **Employer master.csv** – Historical employer account data in the form of a slowly changing

Dimension.

* **Payment transactions.csv** – Superannuation contribution transactions.

**Data Transformation –**

* In Payments transactions.csv file, ‘Cash\_Recieved\_Date’ was transformed into YYYY-MM-DD format to sync with another dataset.
* Column ‘Effective\_To’ in Employer Master.csv file, has end dates of 9999-12-31. These dates were transformed to "2262-01-01" because pandas (Python) do not support this date range and will not impact our metrics anyhow.

**Programming –**

Data Processing (Programming) was done using Jupyter Notebook with Python packages installed.

File Name: Superannaution Metrics 2018.ipynb

Link: <https://github.com/shubhamsakhuja/Superannuation-Metrics-2018>

**Key Insights –**

* Majority of New Employers joined in the month of November 2018 for every Tier (1 - 4) following by the month of December 2018.
* Open Employers by EOM was highest in Tier 4 for all months following by second highest in Tier 3.
* It is worth noticing that in the year 2018, number of Open Employers combined in Tiers 3 and 4 were more than 9 times higher than combined Tiers 1 and 2.
* Total Amount of Payments made by Tier 4 employers were highest following by Tier 2.
* Total number of payments made by Tier 4 employees were the highest.
* However, while calculating metrics it is interesting to see that it takes around 270 payments on average to generate $660122 (appx.) for employers in Tier 4, whereas in Tier 2 it takes around 22 payments on average to generate $480426 (appx.). Hence, it is fair enough to say that employers belong to Tier 2 seems to be contributing the most among all Tiers with respect to total payments made in 2018.

**Final Output –**

Final.csv output generated from Python Script.

Final.xlsx replica of Final.csv with some pivot tables.

Visualisations.pbix – a Microsoft Power BI file for dashboard metrics.

**Future Possibilities – Time Series forecasting and Regression Analysis.**