

Team Meeting 1

01 SEPTEMBER 2022/ 11:13 AM / WORKSHOP

Attendees

Ke, Oliver, Xavier, Tommy, Glendon

Agenda

- ☒ Discuss our approach to the ETL.
- ☒ Plan the next meeting.
- ☒ Plan everyone's next moves.

Notes

- Discussing Tommy's work on the ETL
 - Bugfix reading the whole folder vs reading folder by folder.
 - Addresses fake but postcodes real (for future reference).
- Discussing the Data
 - Surrogate keys: linking user id to transaction user id/consumer id.
 - Check for number of missing consumer ids in consumer table.
 - 3.6 million transactions and Merchant abn == merchant id
- Correlation btn revenue and take rate.
- Discuss need to rework the tag extraction in merchant data (Consider using Pandas).
 - Cleaning tags
 - Capitalizaiton
 - Stemming?
 - OHE?
- Sales by region
 - How do we compare dollar values?
 - How many regions/rows?
- Outlier analysis
 - Is this better kept outside the ETL?

Action Items

- ETL
 - Homogenise dataset fields and how data is read in (Xavier).
 - Extract and link merchant tags to ABS data (Ke)
 - Handling missing values and join dataset (Oliver)
- In depth Analysis:
 - Customer profile by each merchant (gender, postcode, state, orders and dollar spent per time period) (Tommy)
 - Merchant profile: revenue per time period, by region, geospatial visualization, by order (mean, distribution) (Glendon)

Next Meeting

Planned for Monday after 7:00 PM online.