Draft #1

Assumptions:
1. Raw files are in correct
format. No need to check.
2. Use of PostgreSQL database
3. Standardized columns along
L0 and L1 schema.
4. Power BI dashboard will set
for daily sync in case of updates

Interesting datasets:
Link1
Link2

Key Questions:
1. What kind of problem are you
looking for?
Ans. In the area of social media
or social networks.

2. What kind of models are you
looking to apply?
Ans. Online models so it makes
sense to have all the data in the
DB.

3. How does Postgres functions
or ops come into play?

Start

Stores raw data

Raw CSV file → **L0 schema** → **L1 schema** —Postgres ops and ETL→ **L2 schema** → **Power BI dashboard** → **End**

Data aggregation/
Anonymization

Machine Learning