Uber Eats

Predictive Analytics & Dashboarding for ATD Optimization

Miguel Angel Fernandez Castresana

Index

Business context
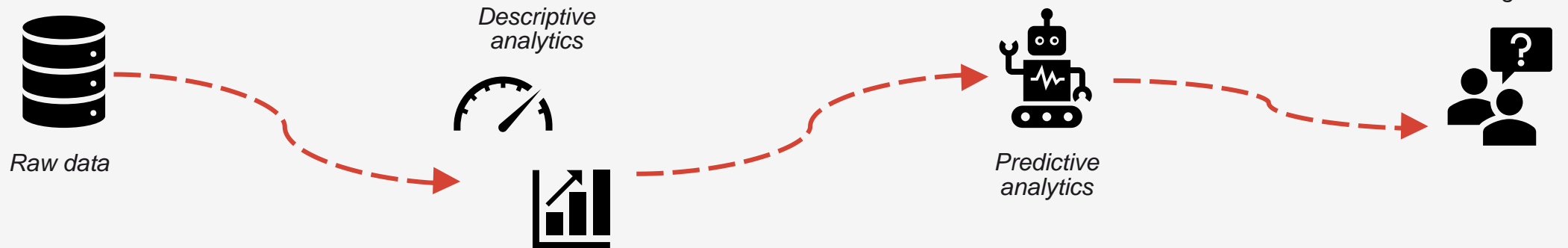*Context and problem statement*

Data	Expected outcome	Need
Comprehensive dataset detailing deliveries across the marketplace in March and April for Mexico	Deliver high-quality descriptive and predictive insights to our customers, empowering them to formulate data-driven strategies; focused in improve the ATD	Develop an automated data extraction pipeline to directly support and enhance descriptive and predictive analytics efforts

Descriptive analytics

Raw data

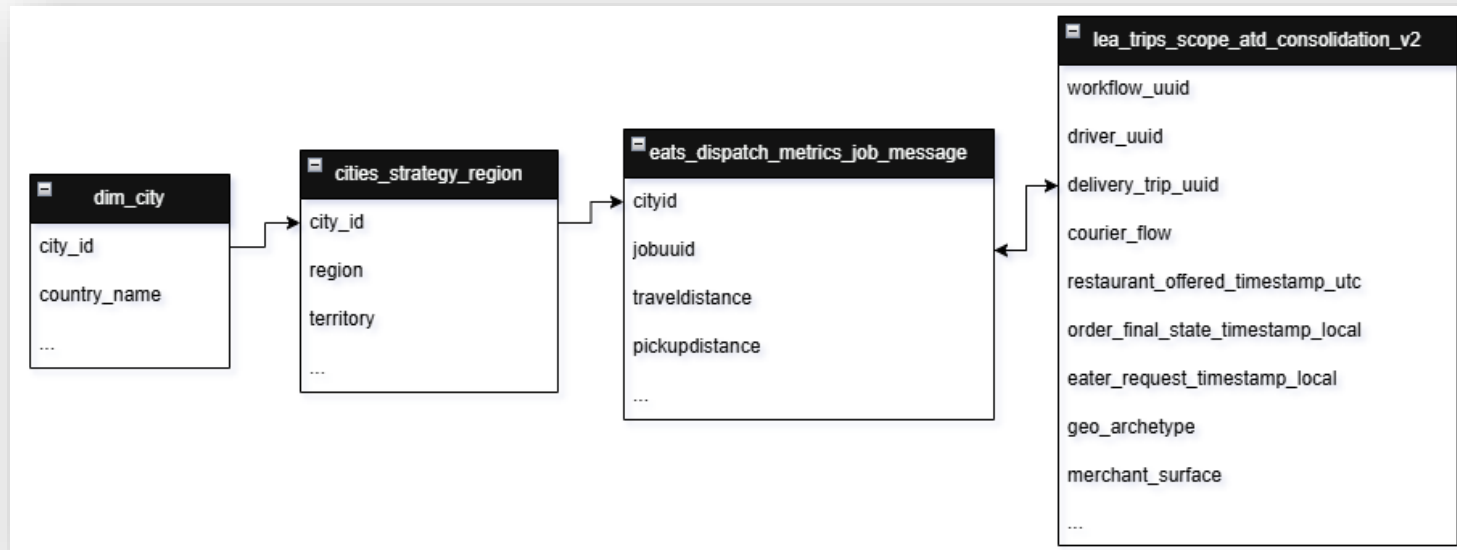*Predictive analytics*

Data-driven strategies

Data extraction
Data model & query composition

Along the query to extract the information, some modifications and adaptations that are important to keep in mind:

- As the date and numerical columns have different format and in some cases null values represented by the text '\N', the query is treating these columns as VARCHAR type and the transformation to the correct format is done along the query

- As in Mexico we have three different hour time zones and we don't have the specification of the state or the *currency_code*; I'm inferring the hours of difference with the *eater_request_timestamp_local* variable

- As asked, for every date provided in the variable *{ds}*, we are providing the deliveries data of the past week, here the calculated week is considered to run from Monday to Sunday
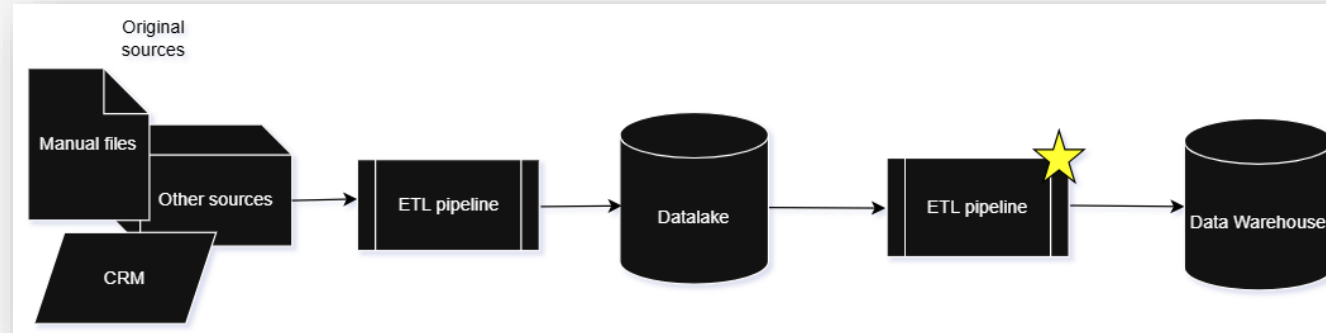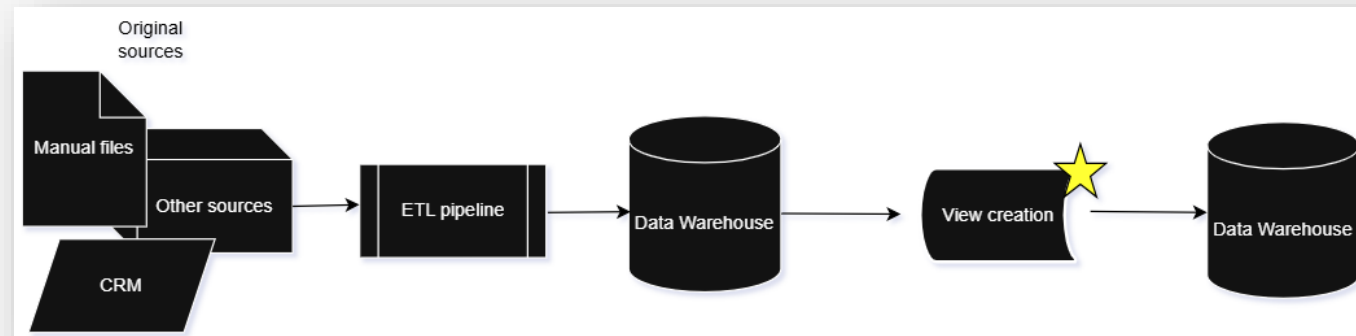
Data extraction
Proposed pipeline

Case 1
The final table where we will storage the data (AA_tables schema) is in a different storage space than the one where the data is coming from, for example the data is coming from the Datalake (Query.sql execution highlighted with a star)

Case 2
The final table where we will storage the data (AA_tables schema) is in the same storage space than the one where the data is coming from (different schemas in the same database) (Query.sql execution highlighted with a star)

Uber Eats

Streamlit dashboard
General composition

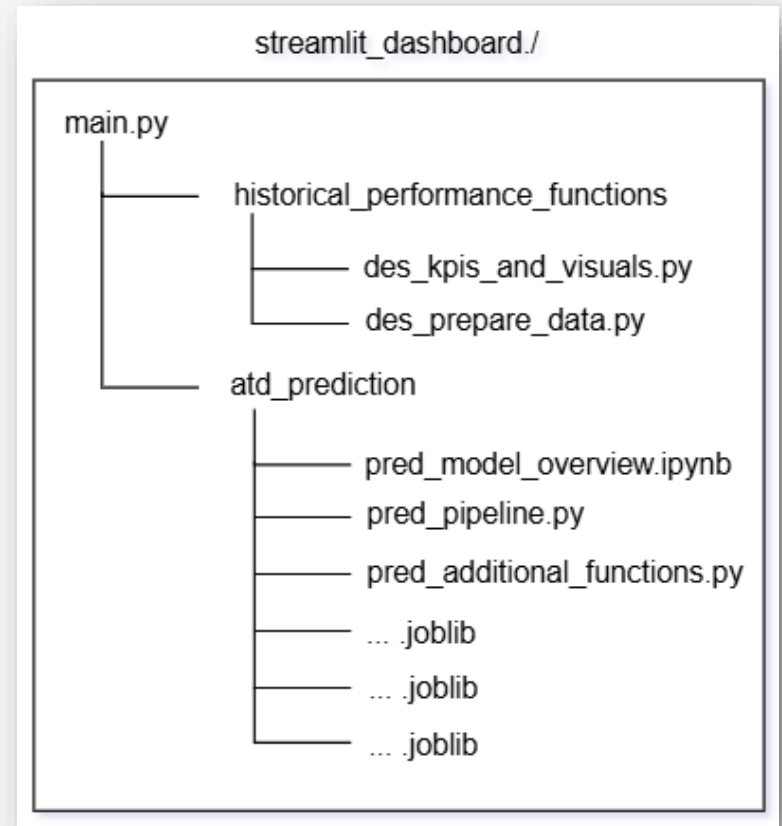## *Composition*

The dashboard is divided in two pages:

Historical & Trend analysis

- Creation of multiple descriptive visualizations to clearly understand data patterns across time, velocity, distance, location, and day of the delivery

- KPIs to track the overall performance

- Development of a time series trend graph integrating SARIMA predictions for the next fourteen days for both the ATD and total delivery volume

ATD prediction

- Analysis of XGBoost feature importance to identify the variables with the highest impact on the ATD prediction

- Implementation and comparison of Random Forest, Multilayer Neural Network (MLNN), and XGBoost models, reporting the specific error metric for each ATD prediction model

Modularization

Streamlit dashboard
General composition

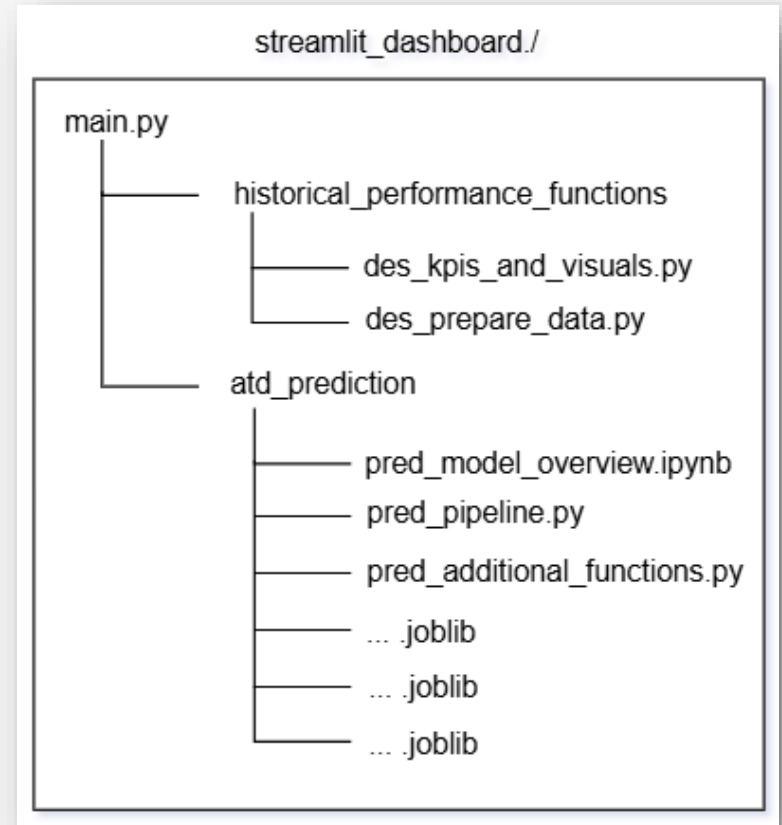## Composition

The dashboard is divided in two pages:

Historical & Trend analysis

- Creation of multiple descriptive visualizations to clearly understand data patterns across time, velocity, distance, location, and day of the delivery

- KPIs to track the overall performance

- Development of a time series trend graph integrating SARIMA predictions for the next fourteen days for both the ATD and total delivery volume

ATD prediction

- Analysis of XGBoost feature importance to identify the variables with the highest impact on the ATD prediction

- Implementation and comparison of Random Forest, Multilayer Neural Network (MLNN), and XGBoost models, reporting the specific error metric for each ATD prediction model

Modularization

```
streamlit_dashboard./

main.py
    ├── historical_performance_functions
    │       ├── des_kpis_and_visuals.py
    │       └── des_prepare_data.py
    └── atd_prediction
            ├── pred_model_overview.ipynb
            ├── pred_pipeline.py
            ├── pred_additional_functions.py
            ├── ... .joblib
            ├── ... .joblib
            └── ... .joblib
```

ATD predictions
ATD variable composition

PENDING

ATD predictions
Missing values, Outliers & Feature Engineering
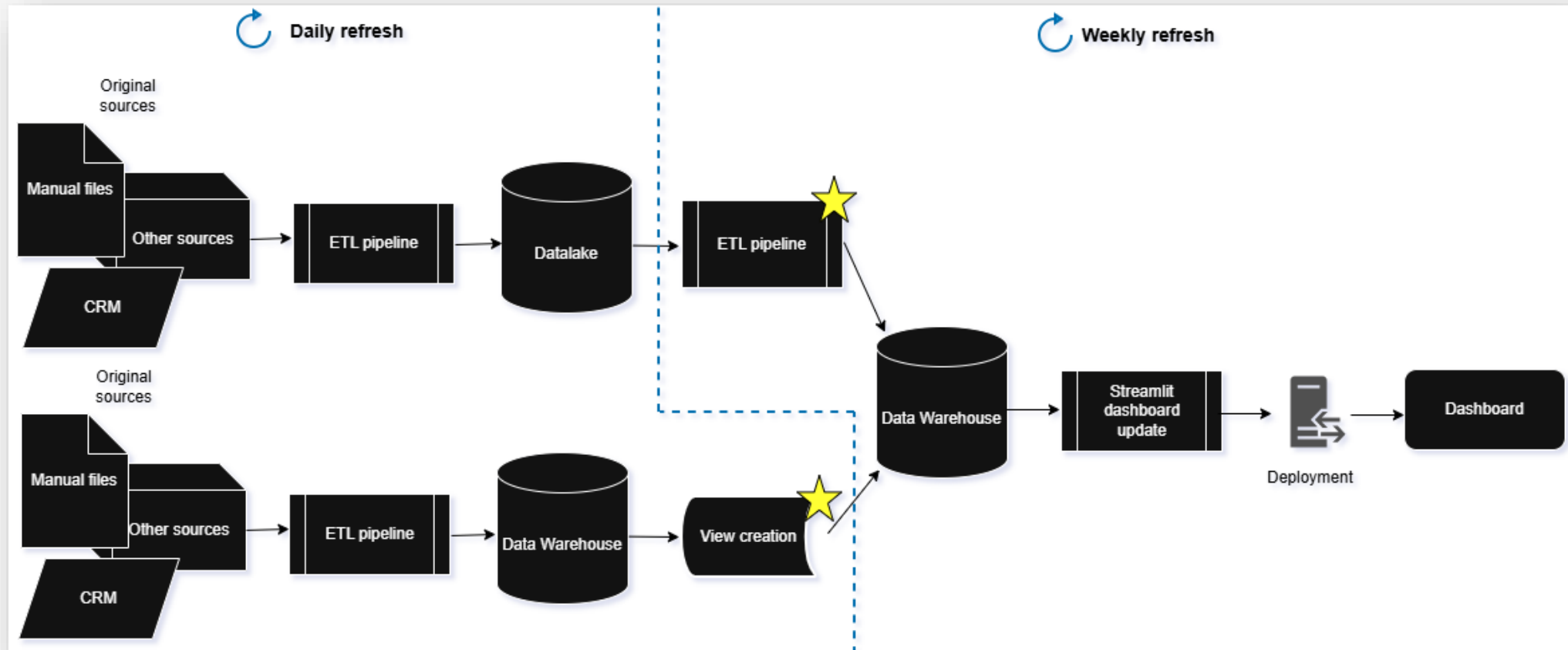
PENDING

Uber
Eats

ATD predictions
Model selection and training

PENDING

ATD predictions
Proposed pipeline

Independent from the case defined for the query update, the proposed pipeline to refresh the dashboard is the following (Query.sql execution highlighted with a star):

This separate pipeline ensures the dashboard's data availability remains unaffected by ongoing query development