Dataflow ETL, ML and Analytics
in Power BI

Daniel Roca

Business Intelligence

dfroca

Data Analytics Enthusiast

Data Scientist, Analytics Consultant, BI Developer

Microsoft Certified: Data Analyst (DA-100)

Analytics roadmap

Introduction

Contents

~~1. Introduction~~

(5 mins)

2. Dataflow ETL

(15 mins)

3. Analytics visuals

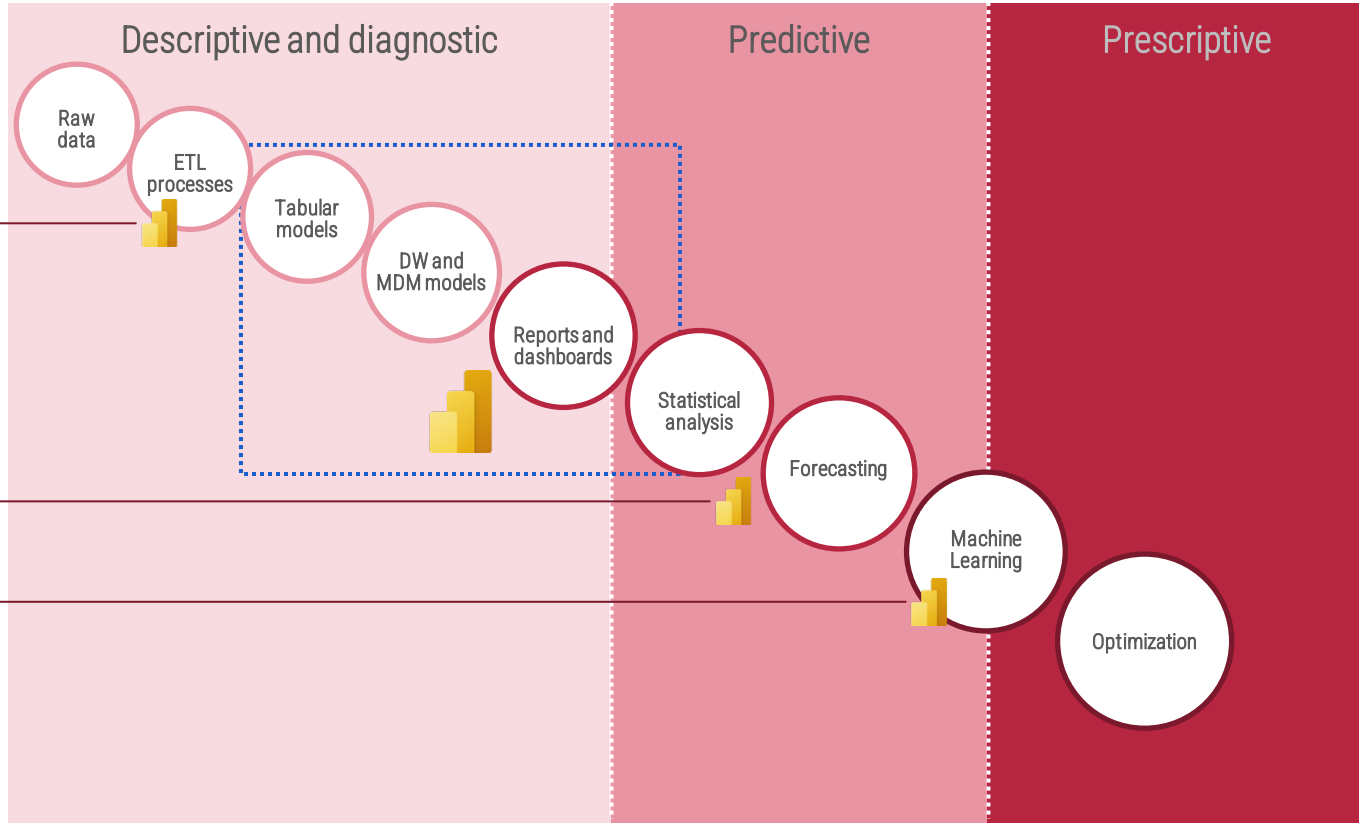
(15 mins)

4. Dataflow Auto-ML

(15 mins)

5. Questions and wrap-up

(5 mins)



Case overview

Introduction

BI team Globant

U.S. Travel Agency

Local flights

Microsoft

Operations based on Azure

Previous year data: airports, airlines, flights **

01

Understand the data

I would like to have the data consolidated and understand the behavior of the flights in the previous year.

02

Delay causes

I want to understand the root causes of flight delays to give better advices to my clients.

03

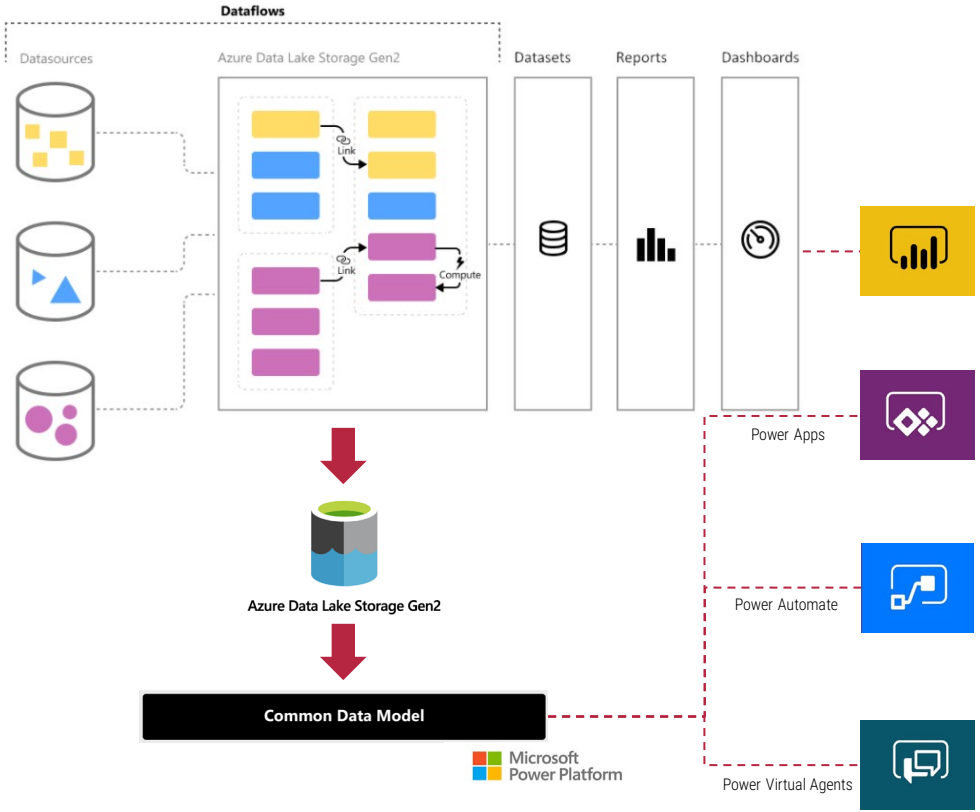
Better assess clients

I'd love to tell my clients which flight should they take based on the minimum probability of delay.

** data based on sample from: kaggle.com/usdot/flight-delays

Key takeaways

Dataflow ETL



Dataflow resulting transformations are stored in files NOT in a database

Very difficult to create a SC Dimensions, Lookups, …

For more efficiency, premium capacity should be enabled

User-friendly transformation tool intended for business users and developers.

Create reusable transformation logic that is used for multiple PBI datasets.

Use resulting data in multiple Power Platform Apps from the CDM

Avoid to overload the Power BI datasets by doing data transformations before, avoiding bottlenecks at a report level

Embedded Auto Machine Learning functionalities within Power BI data flows

Questions and wrap-up

Contents

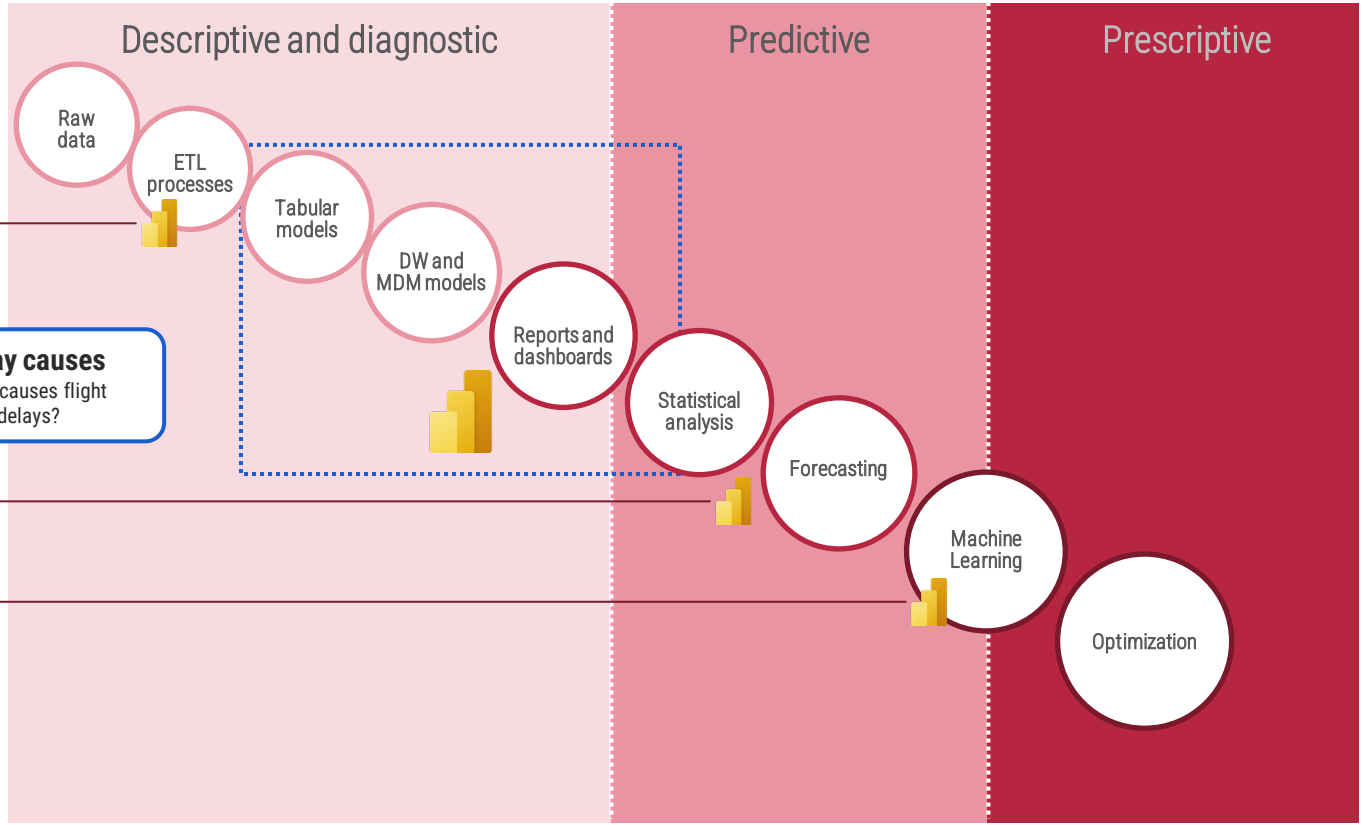
~~1. Introduction~~
(5 mins)

~~2. Dataflow ETL~~
(15 mins)

3. Analytics visuals
(15 mins)

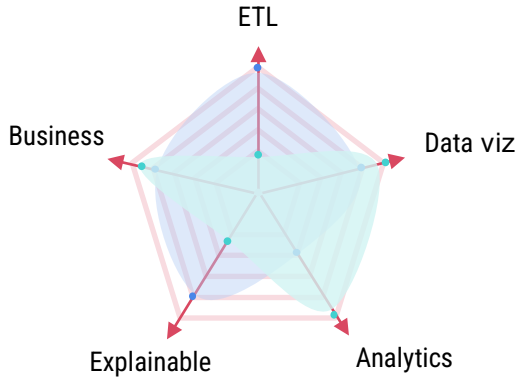
4. Dataflow Auto-ML
(15 mins)

5. Questions and wrap-up
(5 mins)



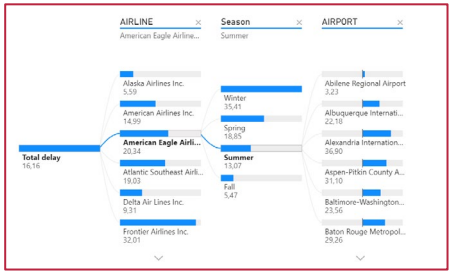
Key takeaways

Analytics visuals



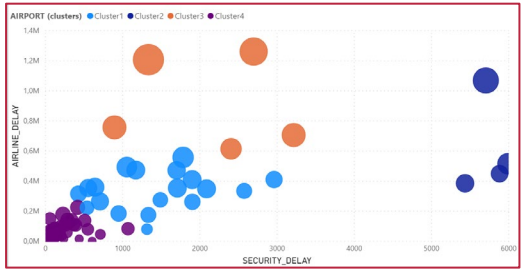
Decomposition tree visual

Statistics – Multivariate analysis



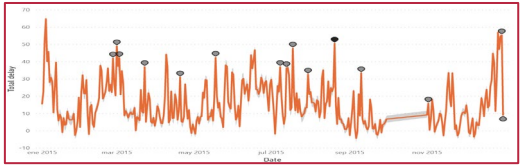
Clusters with a scatter plot

ML – Non supervised learning



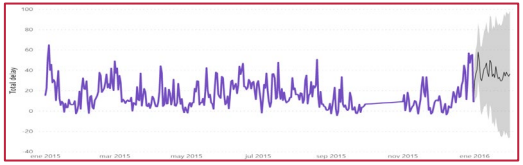
Line chart visual

ML Supervised learning – Time series anomaly detection



Line chart visual

Statistics – Time series forecasting

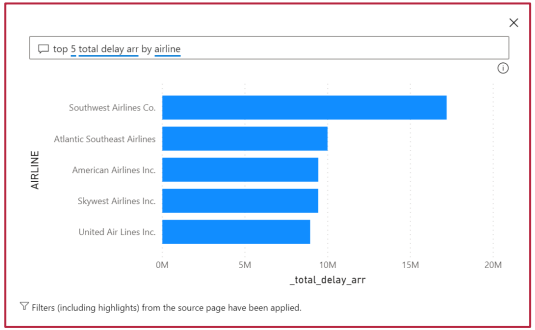


Key influencers visual

ML Supervised learning – Regression algorithms

Q&A visual

AI Text Mining - Synsets



Questions and wrap-up

Contents

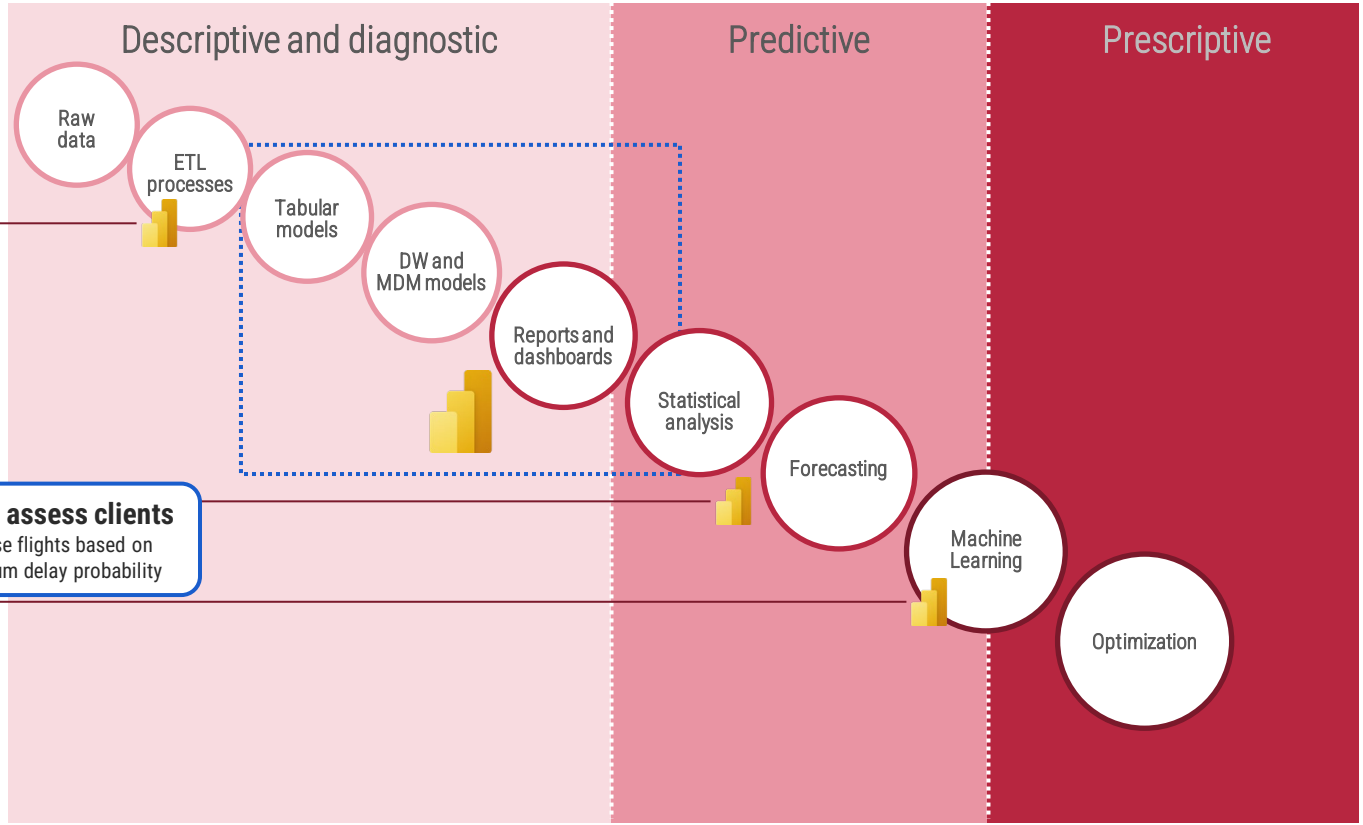
~~1. Introduction~~
(5 mins)

~~2. Dataflow ETL~~
(15 mins)

~~3. Analytics visuals~~
(15 mins)

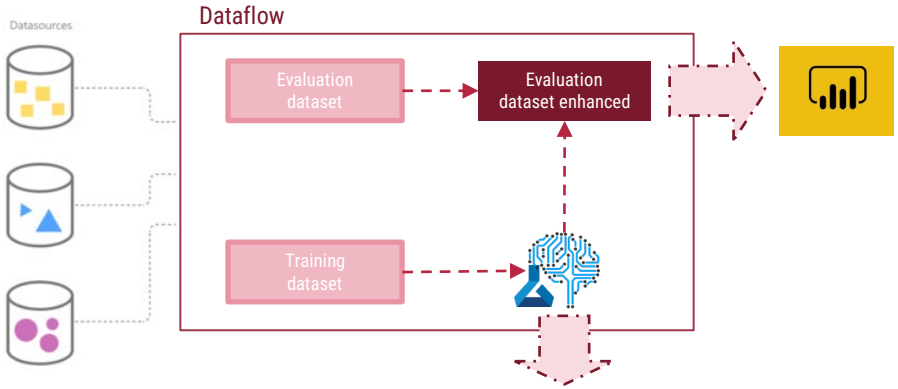
4. Dataflow Auto-ML
(15 mins)

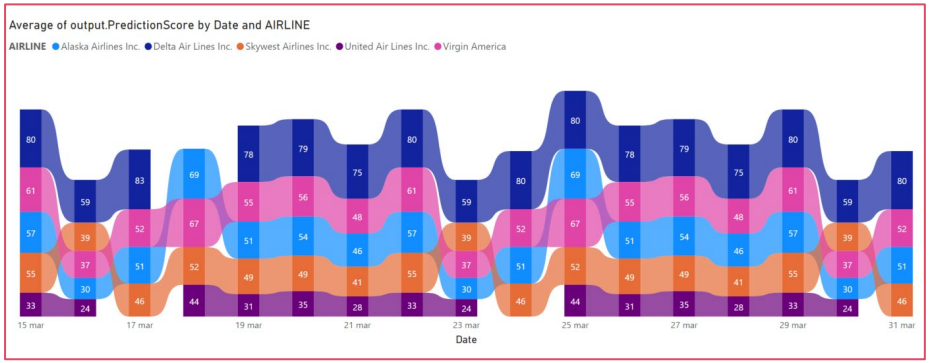
5. Questions and wrap-up
(5 mins)

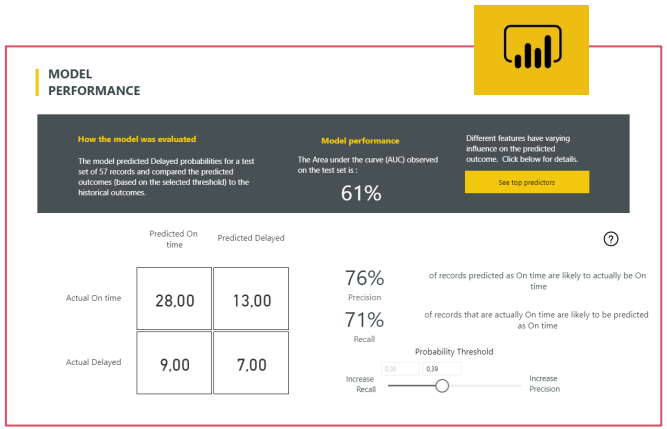


Key takeaways

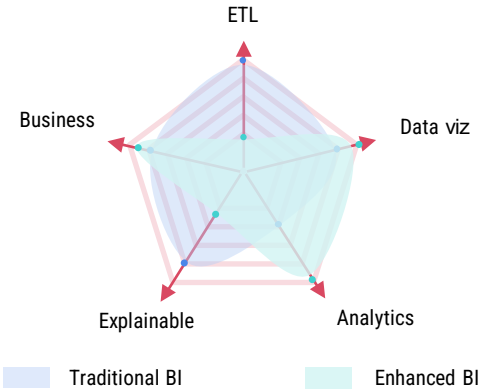
Dataflow Auto-ML







- Auto-ML is available with PBI premium capacity
- ML algorithms for binary classification, multi-class classification and regression tasks.
- Identifies suitable variables to create ML-model (feature engineering)
- Does hyper-parameter tuning, tests various algorithms and provides a report with the training results
- The evaluated dataset is available for use in reports after evaluating the data with the trained ml-model



Questions and wrap-up

Contents

~~1. Introduction~~
(5 mins)

~~2. Dataflow ETL~~
(15 mins)

~~3. Analytics visuals~~
(15 mins)

~~4. Dataflow Auto-ML~~
(15 mins)

5. Questions and wrap-up
(5 mins)