Data Analytics at Gojek

Hanum Kumala
Business Intelligence Analyst
Pasti Ada Jalan
Our Journey

A mobile app for daily needs. The Gojek app offers various services: transport, food delivery, courier logistics, instant shopping, professional massages, payments, and more.

- **2010**: Call-center for ojek services
- **2015**: App launched with 3 services
- **2016**: Expansion and new services, first Unicorn in Indonesia
- **2018**: International growth
- **2019**: Superapp company

*ojek is an Indonesian term of motorcycle ride hailing*
Our Solution for Every Customer’s Needs

goride  
gocar  
gosend  
gobox  
gobluemlbird

gofood  
gobuy  
goshop  
gomed  
gotix

goglam  
gomassage  
godaily  
gomed  
gotix

gopay  
gopoints  
gopulsa  
gobills  
goclean  
goauto
Our Global Footprint

Operating in 207 cities throughout South East Asia

+155m app downloads
+500k merchants
+2m drivers
+60k service providers
Tremendous growth and maturing product lead us to ...

- Bigger quantity and more complex data sources
- Difficulties in getting insight from bunch of data
- Ongoing needs to get real-time insight for business decisions
Business Intelligence at Gojek

- Business Intelligence is the first data team at Gojek

1 of 3 independent data teams besides Data Engineering (focusing on data pipeline in the entire company) and Data Science (focusing on AI and future data products)

- Distributing business data throughout the company: makes data available, aligned with business and accessible by business users

- Thought partner of Product Owners, delivers insights and proves hypothesis based on data
Business Intelligence Process

1. **Data Sources**
   - Historical data transaction

2. **Data Engineering**
   - Pipelines, ETL, EDW, EDL, etc.

3. **Master Data Management**
   - Stewardship, Data Quality, Cleansing

4. **Data Analysis and Reporting**
   - Visualization, Advanced Analytics, KPIs

5. **Decision Making**
   - Deep Knowledge
Data Warehouse Journey
When Gojek need **data at one place?**

We started building the data warehouse from scratch in Q3 of 2016

It started based on the needs of the Business Intelligence team
Labour-intensive work and tons of time to deliver business insights

There was no single place to hold data
Single Version of Truth (SVOT) of business datasets was needed

Various backends across multiple products
Needed to deliver hassle-free, timely and uniform data products
Gojek’s Data Warehouse Architecture

Gojek data architecture as of Q1 2019
Data Quality Service
**Data Quality**

**Completeness**
The proportion of stored data against the potential of “100% Complete”

**Uniqueness**
No thing will be recorded more than once based upon how that thing is identified

**Validity**
Data are valid if conforms to the syntax (format, type, range) of its definition

**Consistency**
The absence of difference, when comparing two or more representation of the metric

**Accuracy**
The degree to which data correctly describes the “real world” object or event

**Timeliness**
The degree to which data represent reality from the required point in time
Data Analysis and Decision Making
Data Analysis Process

Hypothesis or Problem Statements

- Asking the right question for the analysis by defining the hypothesis or problem statements
- What happened?
- Why did it happened?
- What will happen?
- How can I make it happen?
- etc

Get the Data

- Get the data from the Data Warehouse usually using Query
- Data Query
  - Duplication Checking
  - Outlier Checking
  - etc

Data Exploratory

- Exploratory will help us understand what are the things or trends that happened in the data
- Visualization
  - Statistical Analysis
  - Modelling
  - etc

Interpret Result

- Interpreting the data will help to decide whether the hypothesis can be accepted or not
- Actionable Insights
  - Decision Making
  - etc
Product Dashboard

Objective: provide SVOT comprehensive snapshot of business performance

Problem Statements

1. How much has the booking and transaction increased this month?
2. Is there an increasing trend at the festive season?

* Data has been masked due to confidentiality reason
Demand Heatmap

Problem Statements

1. Where does demand come from?
2. Is it permanent or temporary demand?

Main users

- Commercial Expansion
- Merchant Sales

* Data has been masked due to confidentiality reason
Customer Movement

Problem Statements

1. Do our customers stay in the platform?
2. If yes, do they transact often?

Main users

- Strategy
- Product
- Growth

* Data has been masked due to confidentiality reason
Thank you!