
# How Data Science & Analytics Tools are used by the Al Tayyar Group to Support & Enable the Data-Driven Organisation

## DataCon Africa

# Who are we?

Online Business Unit

الطيار
Al Tayyar

Almosafer

tajawal:تجوّل


Vision: To be the **dominant** online travel solution in the **MENA** region, powered by **innovative technology.**

# Almosafer's exceptional Journey of growth

| 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|
| Al Tayyar Group acquires Almosafer | Started partnerships with Al Rahji Bank | **Established Data Team** & expanded into New markets | Crossed the Billion SAR line in Sales | On Track to achieve our goal this year | Determined to achieve our Goal next year |
| **27 Mn** | **74 Mn** | **500 Mn** | **2 Bn** | **3 Bn** | **5 Bn** |

*Sales in SAR*

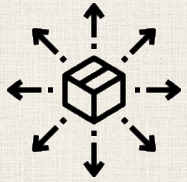
***Largest Online Travel Agency in Saudi Arabia***

# The data strategy follows the customers

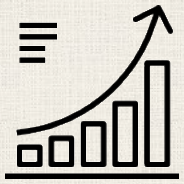
Awareness → Product → Sales → Aftersales

- Awareness: Google AdWords, appfigures
- Product: Google Analytics, adjust, criteo
- Sales: Transactional, Customer Data
- Aftersales: CRM (Customer Relations), 24 (Customer Service)

# Who do we support?



Product



Marketing

Strategy

Finance

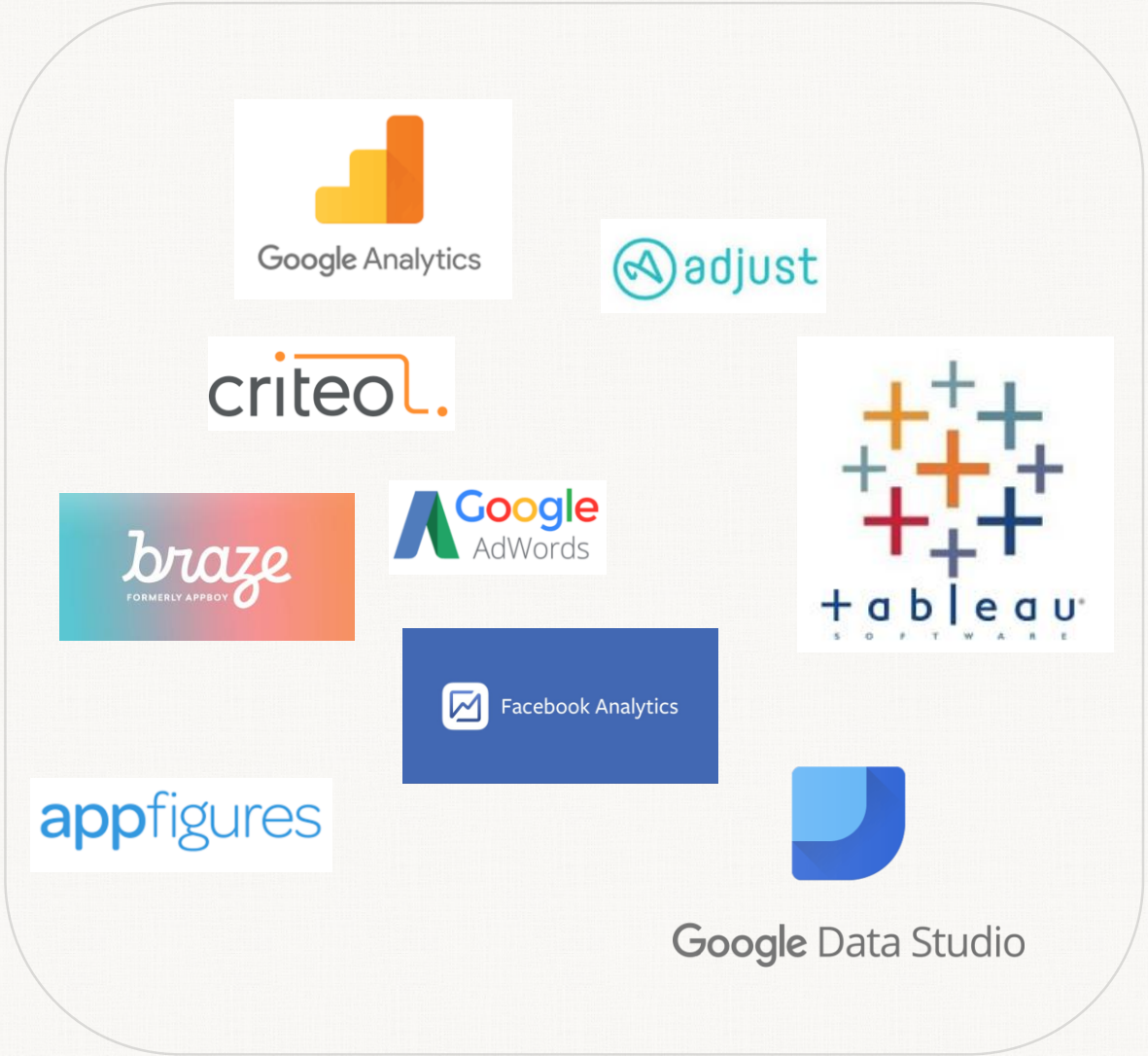
Customer Experience

IT

Commercial

HR

# Different tools for different purposes




Business requirements define the level of reporting and data integration needed.

Tools can be used in isolation or combined for cross-platform analysis where needed.

Implementation of self-service tools are crucial to **empower a data driven organization**

→ Simple requests/tasks can be handled by anyone, leaving more time & resources for deeper business analytics to the Data team.

# How we got started?

...by setting priorities

# I. Cancellation Prediction model v1

Almosafer

tajawal: تجوّل

# Business Background

## Problem statement


- A large share of Hotel bookings made are "Book Now / Pay Later" (~30-50%)
- Depending on seasonality, ~ 70-90% of these bookings are cancelled.
- For 'Pay Now' payment option only ~3% of bookings are cancelled
- Customers are reminded via SMS, in-App, Emails, Phone calls to complete the payment

***"Our revenue estimates for the month are inaccurate due to the large amount of last minute cancelled pay later bookings."***

***"The several follow up attempts require time and effort from our customer service team and cost money, while our customers eventually are inconvenienced."***



Almosafer

tajawal: تجوّل

# AIM OF THE CANCELLATION PREDICTION MODEL

## TO PREDICT THE PROBABILITY OF CANCELLATION OF PAY LATER HOTEL BOOKINGS AT TRANSACTION LEVEL

**USE CASE I.**

Optimizing our process of communication with customers who booked hotels under Pay Later conditions.

**USE CASE II.**

Optimizing marketing spend.

**USE CASE III.**

Incorporating model output in our Revenue forecasting model.

**USE CASE IV.**

Improve accuracy of monthly collections.

# Model Preparation

## Model Selection & Feature Engineering

# Model Selection

We used **Dataiku**, a Data Science platform to run preliminary models on the data and see which algorithm performs best and built our model in Python to implement it on a transaction level


### Random forest

*2018-10-24 – 16:11:07*

| | |
|---|---|
| Trees | **100** |
| Depth | **20** |
| Min samples | **1** |
| Was grid of size | **12** |

***Log Loss:*** *0.093*

### XGBoost

*2018-10-24 – 11:01:32*

| | |
|---|---|
| Trees | **300** |
| Max depth | **5** |
| Was grid of size | **100** |

***Log Loss:*** *0.118*

### Gradient Tree boosting

*2018-10-24 – 05:53:19*

| | |
|---|---|
| Trees | **100** |
| Learning rate | **0.1** |
| Max depth | **3** |

***Log Loss:*** *0.116*

### SVM

*2018-10-24 – 05:47:03*

| | |
|---|---|
| Kernel | **rbf** |
| C | **1.0** |
| Gamma | **0.0** |

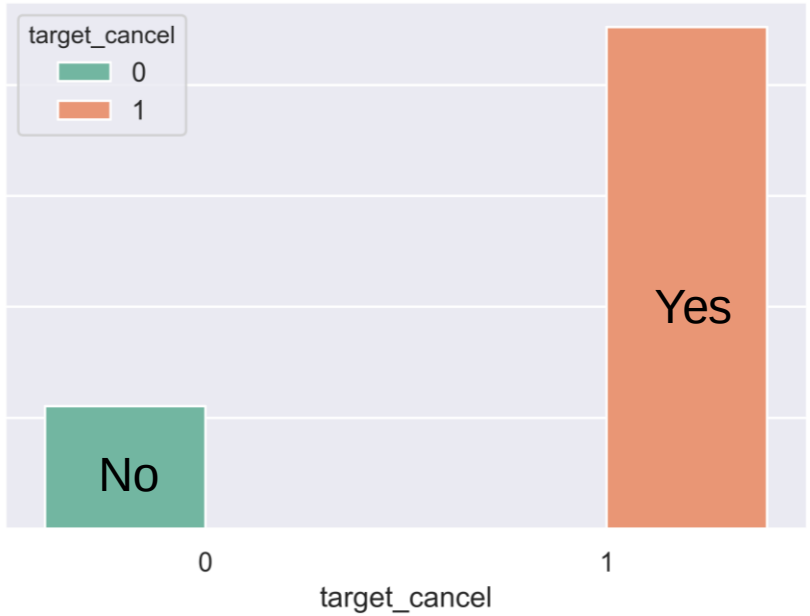
***Log Loss:*** *0.134*

# Feature Engineering

- **Dataset**:
  - transactional data for Hotel and Flight Bookings from Almosafer since 2017 **[ >2,000,000 data points]**
  - Created **34 explanatory variables**, that were potentially expected to influence a customer's cancellation

- Examples:
  - Count of previous cancellations of the customer
  - Customer also booked flight to the same city
  - If customer has any other Hotel booking with the same check-in date
  - Ratio of pay later bookings of each customers
  - Ratio of bookings with promo code per customer
  - CLV vs average historical CLV
  - Average stay duration vs current stay duration per customer

**Distribution of cancelled (1) and not cancelled (0) PAY LATER transactions**



# Model Outcome

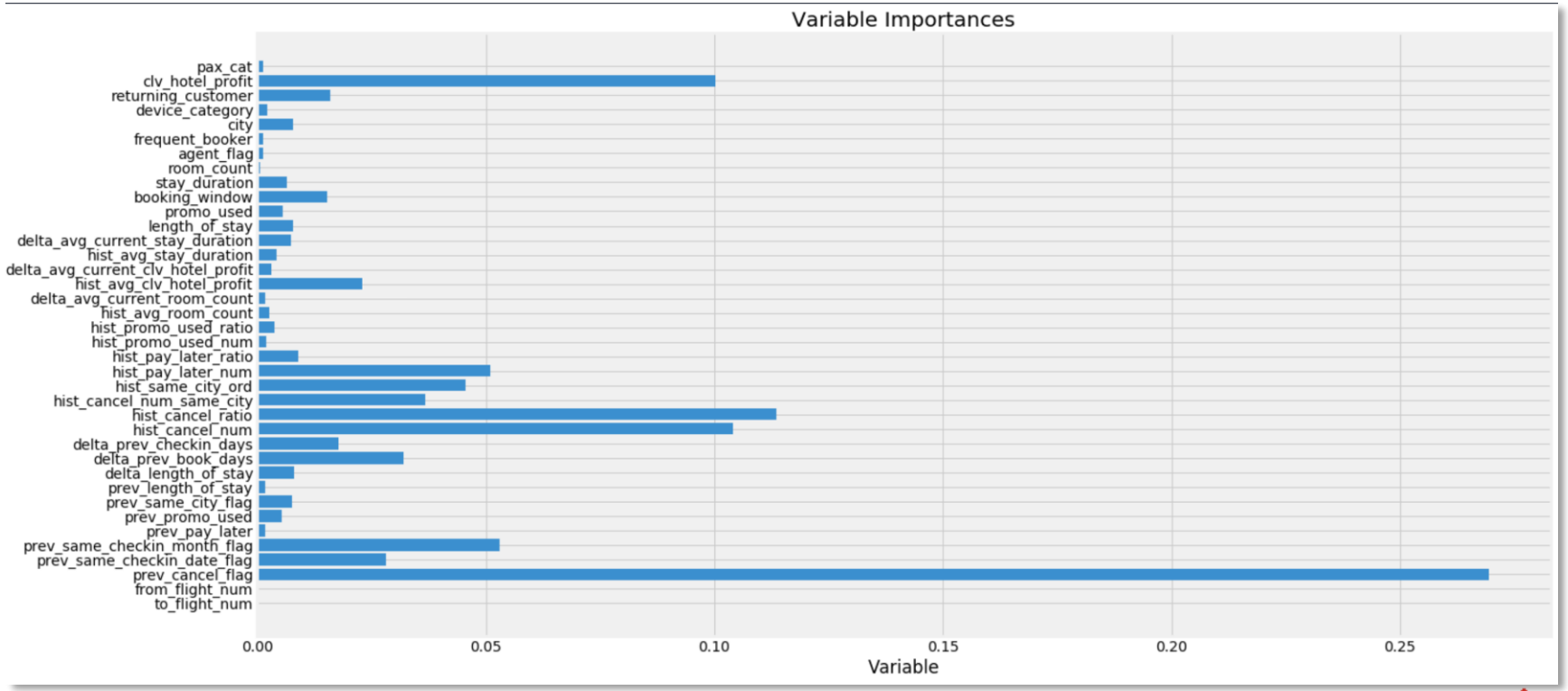
## Model Evaluation & Next Steps

# Model evaluation & Variable Importance

Ongoing optimization of input variables.

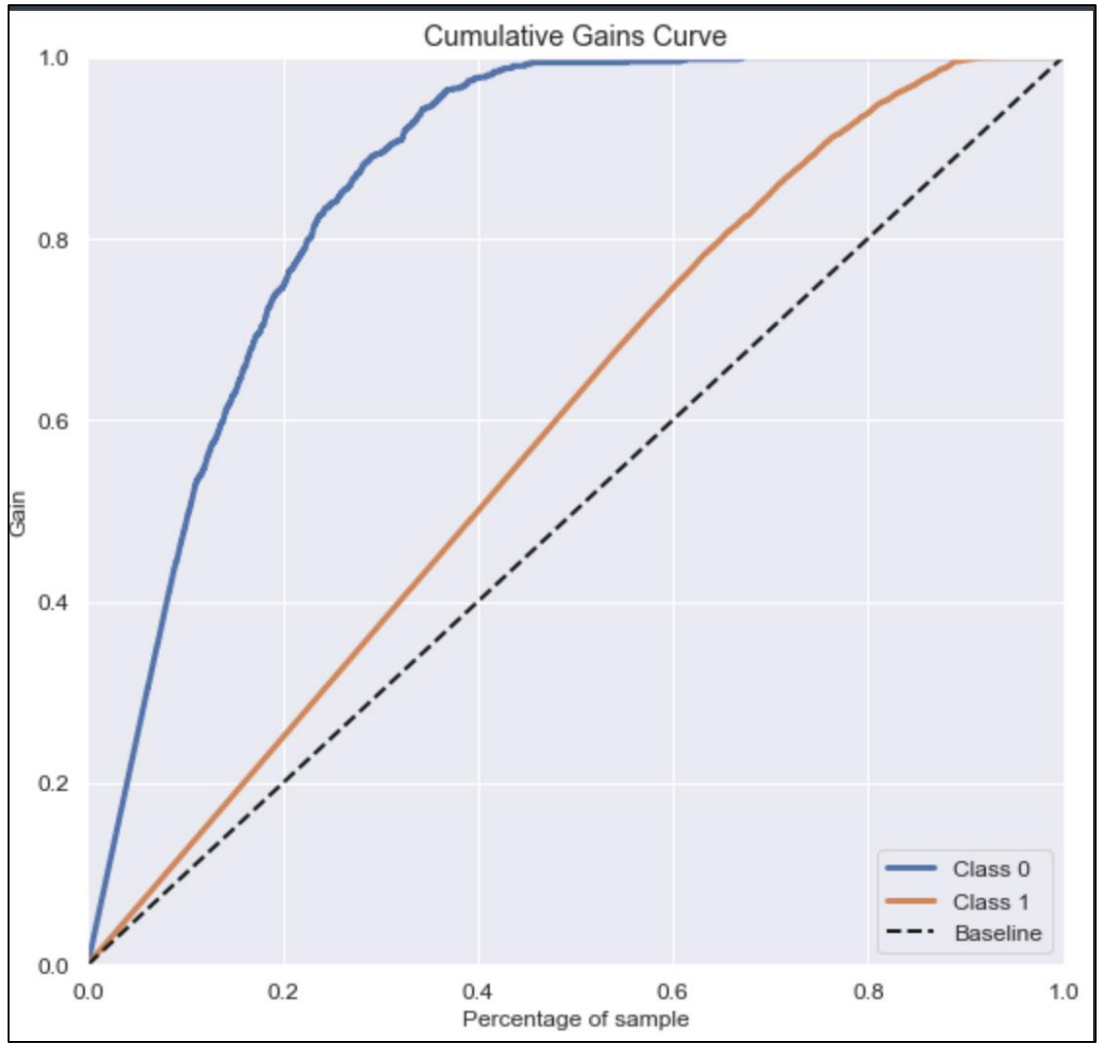
| n = 225,176 | Predicted | |
|---|---|---|
| **Actual** | FALSE | TRUE |
| FALSE | TN = 28,347 | FP = 16,049 |
| TRUE | FN = 4,239 | TP = 176,541 |

- **PRECISION 92%**
- **RECALL 98%**
- **ACCURACY 91%**




- **To be implemented at a transaction level in real time**
- **Model to be enhanced further to improve results even more**

# Cumulative Gain Chart




The goal of this curve is to visualize the benefits of using a model for targeting a subset of the population.

On the horizontal axis, we show the percentage of the population which is targeted and on the vertical axis the percentage of found positive records.

**→This basically represents the benefit of using the model vs without the model.**

The **dotted diagonal** illustrates a random model

# II. BI Chatbot

## Kick-Off v1.0

Almosafer

tajawal: تجوّل

**Problem Statement:**
There a several dashboards & recurring reports that provide rich data for specific purposes.

To support the correct interpretation of this data, a chatbot can be installed on slack to answer common data enquires and enable fast & accurate query resolution.

**Opportunities:**

- Make data more accessible & interpretable
- Provide on instant results for regular requests
- Direct people to appropriate dashboards for further analysis.

# Design Overview

| | Requirements | Possible Approaches |
|---|---|---|
| **Data Layer** | • Access database to be able to respond to requests.<br>• A data-structure to help the bot recommend appropriate tableau dashboards based on the request. | • Specific SQL queries for specific requests.<br>• Read in and store locally the data for specific requests to allow for fast responses. |
| **Business Logic Layer** | • Allow the bot to interpret the message and present the appropriate result for each response. | • Text based response.<br>• Create a small tableau dashboard for each metric and direct people to this for further analysis |
| **Natural Language Processing/ Generative Speech** | • Understand conversational requests from users<br>• Interpret them into an appropriate action/response. | • Procedural/pre-defined responses to key questions.<br>• NLP using deep learning Neural Networks for a more generalised response. |
| **API/ Presentation Layer** | • Send and receive requests/responses to Slack.<br>• Push notifications for recurring updates or "mini-reports" | • The API to communicate with Slack |

**Note:** Tableau provides an API for python. The API only allows you to access the fields/views that are available in tableau. It is not possible to access the data values or filters from the API and therefore this cant be data source.

# NLP – What question is being asked?

This can be used to **determine the context of the question** and therefore the information needed to **derive an appropriate response**.

**1. Procedural keyword identification** and pre-defined responses.

- Straight forward procedural implementation.
- Predictable

**2. NLP**

**A. NLTK** [ http://www.nltk.org/book/]

- The core of most NLP packages in python.
- Very low level and allows you to create very bespoke implementations.
- Can be optimized to be very fast.

**B. SpaCy**

- Wraps the functionality of NLTK into an easier implementation.
- Simplifies many tasks, but doesn't allow for as much freedom as NLTK

# Business Logic – How to respond?

After a question has been identified, this will be used to generate the appropriate response.

1. **Procedural/ structured response** based on the NLP generated keywords.
   - The simplest to implement.
   - Would become very difficult to manage once the business logic become large.
2. **More structured/tree based model**.
   - If used in conjunction with a good classification solution > more manageable than procedural.
   - Would allow for the traversal of the tree for more depth of detail. E.g. revenue -> tajawal -> Organic.
3. **NLP**: Use machine learning to automatically build a response based on previous enquiries
   - Can generate responses based on previous interactions
   - Much more difficult to implement but more scalable.
   - Prone to mistakes if badly trained. Requires good training data (example questions).

→ Need to evaluate Value of each enhancement vs. Effort/Resources

# Putting it together

- **Question:**

"What was the IBV for Almosafer in Q1 2017?"

- **NLP:**

Extract keywords: "IBV", "Almosafer, Q1, 2017.

- **Business Logic:**
  - METRIC [IBV] = Revenue (dictionary Revenue, IBV, Sales, etc..)
  - SUBSET [dim Brand] = Almosafer
  - PERIOD [Year, Quarter] = Year=2017, Quarter=Q1

- **Data Layer:**

Locate the most appropriate response → Revenue_request()

  - Response: The <METRIC> in <PERIOD> for <SUBSET> was <VALUE><FORMAT>
  - Extract Revenue from transactional database for the period and subset and populate the response

- **Response/API:**

Response: "The Revenue in Q1 for Almosafer was $1.57 million"

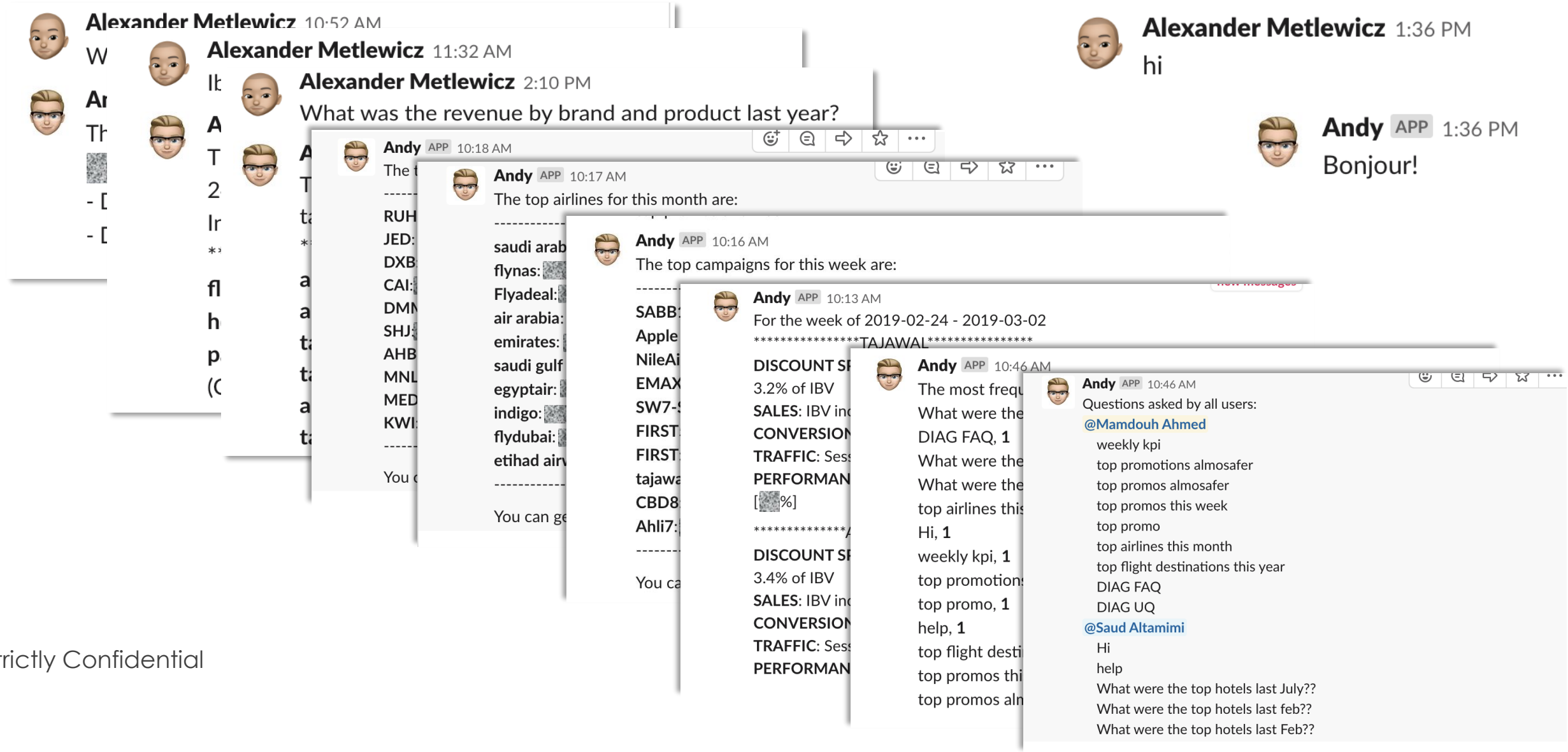
# Scope

- This is an ongoing project that will be completed in phases.
- Each phase will add additional functionality and intellect to the bot.

| | Scope |
|---|---|
| **Phase 1** | • Complete the NLP and API to allow for communication.<br>• Implement basic/common responses in the business logic and data layers. |
| **Phase 2** | • Develop a more saleable/maintainable business logic layer and implement further responses.<br>• Fine tune the NLP layer. |
| **Phase 3** | • Create a business structure for staff:<br>  • Automatically message people with relevant KPI's on a regular basis ("Mini reports")<br>  • Message relevant parties when KPI's are beyond a set limit. (Alerts) |
| **Phase 4** | • Add a more advanced business logic to allow the bot to add helpful details. E.g. revenue is down this week. What might be the cause? |
| **Phase 5** | • Implement a text-to-speech component for audio questions and responses (e.g. alexa/siri integration) |

* Source: Google Analytics (Almosafer Web)

# Andy's responses & reports




# Thank you

Almosafer

tajawal: تجوّل