Ryanair Challenge and Datasets description

1.   Inflight Sales in Ryanair

As the largest airline in Europe, with a fleet of approximately 560 aircrafts, Ryanair operates up to 3,200 flights daily. This extensive operations volume presents significant challenges for the Inflight department, which oversees in-flight sales.

Fresh food is loaded onto aircrafts one time a day prior to the commencement of operations. No additional items are added throughout the day. Each aircraft follows a predefined schedule of flights, known as a line of flight (LoF), which may encompass flights with varying sales patterns. For instance, a short Berlin-Prague-Berlin rotation might be followed by a lengthy leisure rotation from Berlin to the Canary Islands.

Therefore, aircraft connect various flights and destinations along their flight path, where passengers traveling to different destinations may have very distinct characteristics and dietary preferences. Keep in mind that the duration of the flights, their routes (origin and destination), and the characteristics of the passengers will affect their consumption in terms of the number of items, preferences, and so on. The primary challenge for the Inflight department lies in determining the quantity of fresh items to load onto carts, aiming to maximize revenue while minimizing waste.

Due to regular adjustments in the operations department, i.e. mitigating risk of delays, handling technical issues, cancellations, **initially planned lines of flight are not always followed**. This adds extra complexity to the problem as not all the carts follow their originally designed flights.
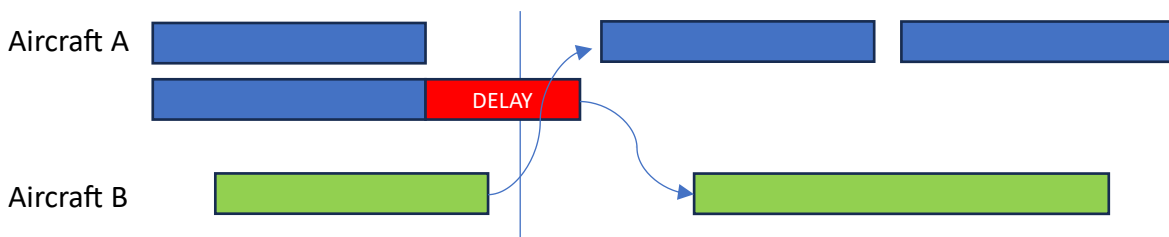
2.   Swap example

An aircraft swap may happen due to various reasons, e.g. moving flights between bases, engineering needs, reducing delays. The chart below presents an synthetic example of a swap related to reducing delay.

Planned

Decision – mitigating reactionary delays

Actual

Aircraft A | DELAYED ROTATION | SWAPPED

Aircraft B | | SWAPPED | SWAPPED

It is important to note that food is not transferred between aircraft. Consequently, when aircraft are swapped, there is a possibility that the food available onboard does not align with the flight's duration, the characteristics, and/or the cultural preferences of the passengers. Keep in mind that each aircraft is stocked with fresh food only once, at the start of its journey (line of flight).

3. Hackathon objective

The objective of the hackathon is to provide recommendations for the Inflight department on how to fill carts with fresh products.

Attendees of the hackathon are given 4 data sets:

1. **Historical sales (train_test)** – Ryanair sales for the full 12m starting from November '22 till Oct '23
2. **Future flights (predict_november)** – details of the November '23 flights. You have to forecast fresh food sales for these flights. Nevertheless, be aware that the output file with your predictions has to contain the forecast at Line of Flights level.
3. **Planned lines of flight (real vs planned lofs_1_year; tab: planned rim lof)** – a list of planned lines of flight from November'22 till November'23.
4. **Actual lines of flight (real vs planned lofs_1_year; tab: real netline lof)** – a list of actual lines of flight from November'22 till November '23.

The goal of this challenge is to recommend loads for **6 fresh product groups per line of flight (LoF)** taking into consideration past patterns to maximize the sales and minimize the waste.

The groups of products sold by Ryanair are as follows:

1. Ham + Cheese Panini
2. Chicken Seeded Panini
3. Ham & Cheese Croissant
4. Fresh Sandwich
5. Choc Croissant
6. Bacon Baguette

These are the 6 fresh product groups that you have to predict for future flights (at LoF level) (File 2 **"predict_november":** containing the list of planned lines of flight for November'23). Be aware that no predictions are made on non-fresh products as they don't deteriorate.

4. Input Datasets

To address this challenge, you have access to three files containing a total of four datasets that you can use to develop your model:

1. **train_test.csv** – historical sales: This file contains all the flights from November' 22 to October'23 including flight characteristics and units sold per product groups, etc. Therefore, the information is provided at the flight level. **It's important to note that a flight line (LoF) consists of several individual flights. Predictions for this challenge have to be made at LoF level.**

COLUMN NAME	DESCRIPTION	DATA TYPE
ST_FLIGHT_ID	Flight code identifier. Concatenation: date + flight number.	string
ST_DEPARTURE_AIRPORT	Departure airport, IATA code	string
ST_ARRIVAL_AIRPORT	Arrival airport, IATA code	string
ST_DEPARTURE_CITY	City of departure	string
ST_ARRIVAL_CITY	City of arrival	string
ST_DEPARTURE_COUNTRY	Country of departure	string
ST_ARRIVAL_COUNTRY	Country of arrival	string
ST_ROUTE_GROUP	Route group, business classification of a given route as defined by the departure and arrival airports	string
ST_ROUTE_SUBGROUP	Route subgroup, sub categories of each route group	string
DT_FLIGHT_DATE	Flight date	datetime
QT_DEPARTURE_LOCAL_TIME	Hour of departure in local time	int
QT_ARRIVAL_LOCAL_TIME	Hour of arrival in local time	int
QT_FLIGHT_DURATION	Flight duration in minutes	int
QT_MONTH_DEPARTURE	Month of departure, month index Jan=1, …, December=12	int
QT_DAYOFWEEK_DEPARTURE	Week day of departure, Sunday=0, …, Saturday=6	int
QT_TOTAL_PASSENGERS	Total number of passengers	int
QT_TOTAL_ADULT	Total number of adults, 16 years old or older	int
QT_TOTAL_TEEN	Total number of teenagers, age between 12 and 15 years old	int
QT_TOTAL_CHILD	Total number of children, age between 2 and 11 years old	int
QT_TOTAL_INFANT	Total number of infants, age below 2 years old	int
QT_TOTAL_MALE	Total number of male passengers	int
QT_TOTAL_FEMALE	Total number of female passengers	int
VL_MEAN_PEOPLE_BOOKING	Average number of people per booking	float

VL_MAX_PEOPLE_BOOKING	Maximum number of people per booking	float
QT_OUTBOUND_PASSENGERS	Total number of outbound passengers	int
QT_RETURN_PASSENGERS	Total number of returning passengers	int
QT_PASSENGER_ONEWAY	Total number of passengers with a one-way flight	int
QT_TOTAL_BAGS	Total number of purchased 10 or 20 Kg bags	int
QT_TOTAL_GOLF	Total number of purchased golf bags	int
QT_TOTAL_SKI	Total number of ski bags	int
QT_TOTAL_BIKE	Total number of purchased bike bags	int
QT_TOTAL_BULK	Total number of other purchased items to carry, such as baby equipment	int
QT_TOTAL_MUSIC	Total number of purchased instruments to carry	int
QT_TOTAL_SPORTOTHER	Total number of sport baggage other than the ones already mentioned	int
QT_TOTAL_PLUS	Total number of purchased plus bundle units	int
QT_TOTAL_FLEXIPLUS	Total number of purchased flexi plus bundles units	int
QT_TOTAL_BOOKINGS	Total number of bookings	int
QT_TOTAL_1MALE	Total number of bookings with a single male passenger	int
QT_TOTAL_1FEMALE	Total number of bookings with a single female passenger	int
QT_TOTAL_2MALES	Total number of bookings with two male passengers	int
QT_TOTAL_2FEMALES	Total number of bookings with two female passengers	int
QT_TOTAL_3MALES_PLUS	Total number of bookings with three or more male passengers	int
QT_TOTAL_3FEMALES_PLUS	Total number of bookings with three or more female passengers	int
QT_TOTAL_3MALEFEMALE_MIXED	Total number of bookings with three passengers of mixed genders	int
QT_TOTAL_REST_COMPOSITION	Total number of bookings with other gender compositions	int
DAY_OF_ORIGIN	Departure day	datetime
DEP_SCHED_DT	Scheduled departing time	datetime
ARR_SCHED_DT	Scheduled arrival time	datetime
FN_CARRIER	Carrier code	string
AC_LOGICAL_NO	Aircraft number	int
BASE	Planning base where the aircraft is loaded, IATA code	string

LOF_ID	Identifier for a line-of-flight, set of airports connected with the same aircraft within a day	int
FN_NUMBER	Flight number	int
FLIGHT_N	Flight number with carrier	string
DEP_DELAY	Departure delay in minutes, when positive this means that the departures is postponed from its scheduled departure time and vice versa	int
HAM_CHEESE_PANINI_QT	Sold units of Ham & Cheese panini	int
CHICKEN_SEEDED_PANINI_QT	Sold units of Chicken Seeded panini	int
HAM_CHEESE_CROISSANT_QT	Sold units of Ham & Cheese croissants	int
FRESH_SANDWICH_QT	Sold units of Fresh sandwiches	int
CHOC_CROISSANT_QT	Sold units of Chocolate croissants	int
BACON_BAGUETTE_QT	Sold units of Bacon baguettes	int

2. predict_november.csv

The file contains a list of flights from November 2023, each accompanied by their respective attributes. Your task is to forecast the purchase of fresh food for these specific flights. The file's format mirrors that of the 'train_test.csv' file; however, it does not include actual data on fresh food sales, which is the data you are required to forecast. **While the provided file details individual flights, the challenge lies in predicting sales of fresh products at the Line of Flight (LoF) level.** This will necessitate engaging in various aggregation and drill-down exercises.

3. real vs planned lofs_1_year.xlsx. This file contains two tabs:

a) RIM lof – planned lines of flight since November 2022 till November 2023.

This tab provides information about the **planned flight of lines (LoF) and each of its flights**. However, this planning may undergo changes due to business needs or unforeseen events. Therefore, information is provided at LoF level.

COLUMN NAME	DESCRIPTION	DATA TYPE
DATEOFORIGIN	Planned day of operations	date
AIRPORTBASE	Aircraft base, IATA code	string
FLIGHTS	List of flight numbers	string
HOPS	Planned line of flight. For instance AGA / BGY / AGA / MRS / AGA means that the aircraft is flight 4 flights on a given day: AGA – BGY, BGY – AGA, AGA – MRS, MRS - AGA	string

b) Netline actual lof since November 2022 till November 2023.

This tab offers information about the **actual flight of line (LoF) and each of its flights**. Therefore, information is provided at LoF level. Considering the differences between planned

and actual line of flights or flights can assist you in making more accurate predictions if certain patterns are identified.

Please note that you can cross-reference this information with historical flights by using the flight number. The flight number in this file can be matched with the flight code identifier in the historical flights file, which consists of a concatenation of the date and the flight number. In addition to this, remember that predictions have to be made at LoF level.

COLUMN NAME	DESCRIPTION	DATA TYPE
BASE	Aircraft base, IATA code	string
LOF_ID	Line of flight identifier	string
DATE_OF_ORIGIN	Planned day of operations	date
FLIGHTS	List of flight numbers	string
HOPS	Actual line of flight. For instance AGA / BGY / AGA / MRS / AGA means that the aircraft is flight 4 flights on a given day: AGA – BGY, BGY – AGA, AGA – MRS, MRS - AGA	string

5. Technical materials to submit:
Here you can find the technical materials to submit:
(a) The code they developed to address the problem (ideally with comments, or Jupyter notebooks alternating between code and text explanations).
(b) The output file containing your predictions (by Google Forms).
(c) The model performance metrics (by Google Forms).

o   **Output file with your predictions:**
You are required to develop a model that predicts sales for six fresh product groups for each Line of Flight (LoF) corresponding to the flights listed in 'Predict_november.csv'. The forecasts should be made at Line of Flight level and for each of the six fresh product groups. You are allowed to employ innovative methods to enhance the robustness of your predictions, such as experimenting with different methods and levels of aggregation. You are encouraged to be as creative as you wish, while upholding technical and analytical precision. However, keep in mind that the performance metrics must be calculated for the predictions at the Line of flight, considering all six fresh product groups simultaneously.

Note that the forecasted quantities of the products must be integers (they cannot be split into pieces).

The output dataset, which should contain your forecast for fresh product group sales for flights in November 2023, must be structured as follows and saved in a CSV format:

COLUM NAME	DESCRIPTION	DATA TYPE
LOF_ID	Line of Flight Identification.	string
HAM_CHEESE_PANINI_QT	Forecasted units of Ham & Cheese panini	int

CHICKEN_SEEDED_PANINI_QT	Forecasted units of Chicken Seeded panini	Int
HAM_CHEESE_CROISSANT_QT	Forecasted units of Ham & Cheese croissants	Int
FRESH_SANDWICH_QT	Forecasted units of Fresh sandwiches	Int
CHOC_CROISSANT_QT	Forecasted units of Chocolate croissants	Int
BACON_BAGUETTE_QT	Forecasted units of Bacon baguettes	int

o   **The model performance metrics.**

You are required to provide the following performance metrics, each with **at least four decimal places**:

1. Mean Absolute Percentage Error (MAPE)
2. Root Mean Square Error (RMSE)
3. Mean Absolute Error (MAE)

The superior predictive model will be determined by the smallest values across all three metrics. These performance metrics must be calculated for predictions at the line of flights level, taking into account all six fresh product groups simultaneously.

These **performance metrics (MAPE, RMSE and MAE)** and the **output file with your predictions** have to be submitted to a **Google Forms** following the instructions.
Please note: Access is restricted to those using their IE email account.

○ **How to compare your performance metrics with other Teams:**

You can check your performance metrics compared with other teams accessing this **Google Forms** anytime (See the section highlighted in yellow in the image below)
Please note: Access is restricted to those using their IE email account.

IE Sustainability Datathon 2024

February 8th - March 6th, 2024

IE Datathon 2024

1. MODEL EVALUATION CRITERIA:

Submissions will be evaluated employing three selected metrics. *The model demonstrating superior predictive capability will display the smallest values across all three metrics.*

1. Mean Absolute Percentage Error (**MAPE**)
2. Root Mean Square Error (**RMSE**)
3. Mean Absolute Error (**MAE**).

2. FORMAT:

Ensure that the metrics are presented to <u>AT LEAST FOUR DECIMAL PLACES</u>. *Submissions that do not adhere to the specified format will not be considered for competition.*

Example Formats:

1. MAPE (%): 8.6893
2. RMSE: 0.5939
3. MAE: 0.4984

3. OUTPUT FILE FORMAT (CSV):

The output dataset, which should contain your forecast for fresh product group sales for **Line of Flight in November 2023**, must be structured as follows and saved in a CSV format:

- **LOF_id**: Flight code identifier. Date + Flight number. (String)
- **Ham_Cheese_Panini_qt**: Forecasted units of Ham & Cheese panini (integer)
- **Chicken_Seeded_Panini_qt**: Forecasted units of Chicken Seeded panini (integer)
- **Ham_Cheese_Croissant_qt**: Forecasted units of Ham & Cheese croissants (integer)
- **Fresh_Sandwich_qt**: Forecasted units of Fresh sandwiches (integer)
- **Choc_Croissant_qt**: Forecasted units of Chocolate croissants (integer)
- **Bacon_Baquette_qt**: Forecasted units of Bacon baquettes (integer)

Note that the forecasted quantities of the products must be integers (they cannot be split into pieces).

4. CHECK YOUR RESULTS:
LINK TO RANKINGS