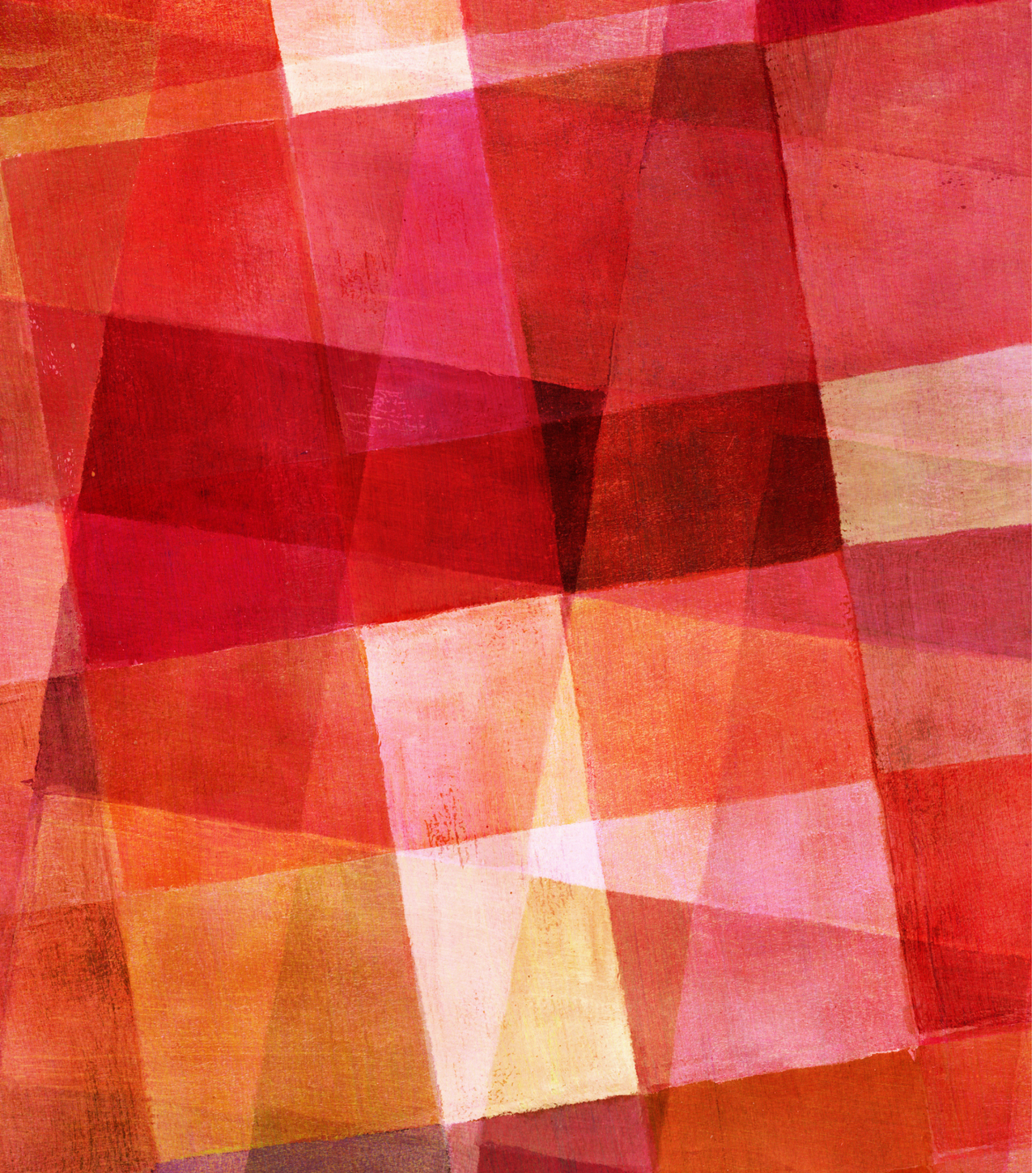
# A ML journey from customer reviews to business insights

*Dr. Federica Lionetto*

UZH ML Workshop - 17 November 2020



# AGENDA

First part: 14:00-14:45

- Introduction of the use case
- Key information on the dataset
- Data preparation and exploratory data analysis

Coffee break: 14:45-15:00

Second part: 15:00-15:45

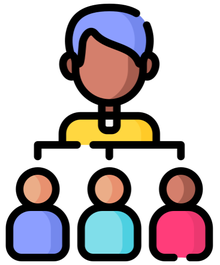
- Modelling
  - training and test
  - performance evaluation
  - black box vs. model explainability
- Word clouds as a way to visualise results

Q&A: 15:45-16:00


# INTRODUCTION OF THE USE CASE

# WHY CUSTOMER REVIEWS?

- **Customer reviews** are almost ubiquitous, and for a good reason: they help both customers and product/service providers to set and reach high standards for customer experience.
- The **value**: The ability to **promptly** and **regularly** understand customers' satisfaction and its key drivers can provide a competitive advantage to a company. In particular, it allows to:
  - inform strategies for customer acquisition and retention
  - trigger remedial actions to prevent customer churn
  - highlight the most promising R&D areas within the company
  - identify opportunities for new or better products/services
  - personalise the customer experience



- The **challenge**: Extracting business insights from customer reviews is time consuming and hardly manageable through a manual process.
- The **solution**: ML and NLP can speed up the process by automating the algorithmic and repetitive part of the workflow.

PART 2.

# KEY INFORMATION ON THE DATASET

# WHICH DATA?

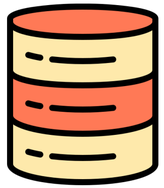
- We will consider a real-world use case: **airline customer reviews**.
- The dataset is scraped from Skytrax and is publicly available at: https://www.kaggle.com/efehandanisman/skytrax-airline-reviews

10/10 **"Can't fault their customer service"**

Gregory Epps (South Africa) 5th October 2020

✅ *Trip Verified* | Can't fault their customer service. My wife's original flight was cancelled, I called them up and had it rebooked in under 3 minutes and got the date she wanted. No hassle. Happy customer!

| | |
|---|---|
| **Type Of Traveller** | Solo Leisure |
| **Seat Type** | First Class |
| **Route** | Frankfurt to Johannesburg |
| **Date Flown** | October 2020 |
| **Seat Comfort** | ★★★★☆ |
| **Cabin Staff Service** | ★★★★★ |
| **Ground Service** | ★★★★☆ |
| **Value For Money** | ★★★★☆ |
| **Recommended** | ✔ |



*>130k records*

*17 fields*

*verified customer reviews submitted between 2002 and 2019*

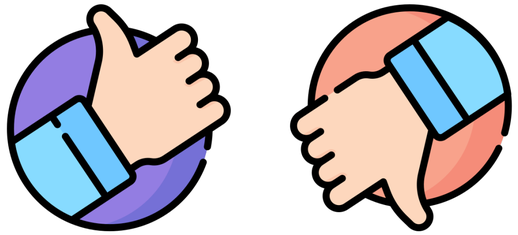


Icons designed by Freepick from Flaticon

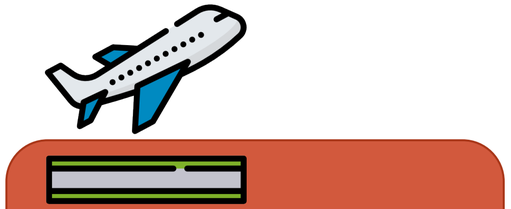

# TODAY'S TUTORIAL

# WHAT ARE WE GOING TO MODEL?

- The main goal of today's tutorial is to familiarise ourselves with some of the many interesting tools for ML and NLP.
- In order to do that, we will set a practical objective, that is, to train a ML model that can predict whether a customer review is positive or negative, that is, if the customer is recommending the service to others.
  - We can frame this as a **binary classification** problem to solve with a **supervised learning** approach.
  - The **label** is represented by the yes/no value of the "recommended" field.
  - The input features are those available in the initial dataset, augmented through **feature engineering**.



# OUR TUTORIAL STEP BY STEP



Customer reviews

Business insights

- Data gathering and exploratory data analysis
- Data cleaning and preprocessing
- Feature engineering
- Model development
- Performance evaluation
- Interpretation of the predictions

# WHAT TO EXPECT FROM TODAY'S TUTORIAL

# Thank you for your attention…
# …and let's stay in touch!



*federica.lionetto@gmail.com*


*@federica-lionetto*


Icons designed by Freepick from Flaticon