### Documentation

## motivation

*# Motivation for data context*

The data scraped in this report will give assess to insides about KLM Royal Dutch Airlines customer reviews places on the website airlinequality.com. The scraping code that has been used to scrape the data from KLM’s page can also be used to scrape data from other airlines. After scarping the reviews of KLM, a dataset has been made of the information that has been scraped.

According to research of the European.ce (2020) 1,034 million Europeans travelled by plane. This is an increase of 3.8% compared with 2018. Corona has shut down the aircraft industry for a while, but the industry is currently getting back on track. Because this market has so many customers, there are also many suppliers. It is estimated that there are around 5.000-5.500 airlines across the world, of which there are 700-800 commercial airlines operating scheduled flights globally.

According to research airlines have to focus on customer satisfaction and customer experience to differentiate itself from other competitors in this market (Ban & Kim, 2019). Studies have shown that analysing online review data has several advantages for the customers’ satisfaction and the way the customers experience the airlines. These advantages are:

* It is an inexpensive way to gather information as reviews are often already available on websites or apps (Liau & Tan, 2014).
* Reviews from customers are considered as trustworthy (Brochado et. Al, 2019).
* Reviews are fast, which means that reviews are often put online within days of a customer's "purchase" (Brochado et. Al, 2019).

These advantages show that gaining data from reviews of customers can be an interesting way to analyse and improve customer satisfaction. However, there is no dataset available where airlines can analyse all reviews of their airline. Analysing reviews from website is extremely time consuming because little to no filters can be applied and therefore it is very difficult to make segments. Scraping the reviews from a website to make a dataset, can therefore be very useful for airlines.

# Motivation for website

There are plenty of websites on the internet where reviews of airlines are collected. Several websites have been analysed and the website airlinequality.com has been chosen as the best option for scraping because of the following reasons:

* Airlinequality.com gives customers the opportunity to verify their flight though by uploading their boading pass or ticket. Airlines can therefore be sure that the verified reviews are written by people who have actually used the airline. As a result, the dataset contains more credible data.
* Airlinequality.com lets customers rate the airline on several different variables which already have been mentioned above. This gives the airline star ratings on various variables which gives the airline insight into its good and bad points.
* Airlinequality.com is an international website, which means that it can get reviews from customers of all countries.

The websites that have been analysed besides airlinequality.com are mentioned below and it is given why these website will not be used to scrape reviews.

* Tripadvisor.com. This website lets the customer rates on different variables but does not give verified reviews, which means that the website does not check whether the customer has actually been on the flight he is reviewing. This allows people to write fake reviews, this reduces the reliability of the dataset.
* Flight.report.com. This website lets customer also rate on different variables but each review is published on a separate page. When scraping these reviews, each page should be scraped individually to get all the reviews. This is extremely time consuming and therefore these website has not been chosen to scrape.

# *Entities, linkages, timeframe and algorithmic biases*

The dataset from

Algorithmic biases??

The dataset contains the following information:

* KLM's average number of stars on the variables: Food & Beverages, Inflight Entertainment, Seat Comfort, Staff Service, Value for Money.
* Full reviews of customers including:

|  |  |
| --- | --- |
| the name of the customer | Character |
| Land of origin | Character |
| Date of the review | Date |
| Date of flight | Date |
| Type of aircraft | Character |
| Type of Traveller’ | Character |
| Seat type | Character |
| Route of the flight | Character |
| Rating of seat comfort, Cabin Staff Service, Ground Service, Value for money, Food & beverages, Cabin WiFi & connectivity andinflight entertainment | Numeric |
| Whether the customer would recommend the airline | Logical |
| A written review of their experience | Character |

# *Entities, linkages, timeframe and algorithmic biases*

[Yee Liau, B.](https://www.emerald.com/insight/search?q=Bee%20Yee%20Liau) and [Pei Tan, P.](https://www.emerald.com/insight/search?q=Pei%20Pei%20Tan) (2014), "Gaining customer knowledge in low cost airlines through text mining", [*Industrial Management & Data Systems*](https://www.emerald.com/insight/publication/issn/0263-5577), Vol. 114 No. 9, pp. 1344-1359. <https://doi.org/10.1108/IMDS-07-2014-0225>

[Brochado, A.](https://www.emerald.com/insight/search?q=Ana%20Brochado), [Rita, P.](https://www.emerald.com/insight/search?q=Paulo%20Rita), [Oliveira, C.](https://www.emerald.com/insight/search?q=Cristina%20Oliveira) and [Oliveira, F.](https://www.emerald.com/insight/search?q=Fernando%20Oliveira) (2019), "Airline passengers’ perceptions of service quality: themes in online reviews", [*International Journal of Contemporary Hospitality Management*](https://www.emerald.com/insight/publication/issn/0959-6119), Vol. 31 No. 2, pp. 855-873. <https://doi.org/10.1108/IJCHM-09-2017-0572>

a few recent studies have highlighted the advantages of analyzing online review data for studying customers’ satisfaction or their experience of the airline

When choosing an airline, a choice often has to be made between price and service. Where one wants to fly cheaply, another finds service more important. But it is expected that trust is also a criteria that many passengers take into consideration since everyone wants that the airline brings you safely from one place to another. That is why we expect that reviews, for example, are very important in this sector and useful to analyse.

**n 2019, 1 034 million people in the EU travelled by air, an increase of 3.8 % compared with 2018.** According to research from CBS .. % of the Europeans uses a plane at least once a year.