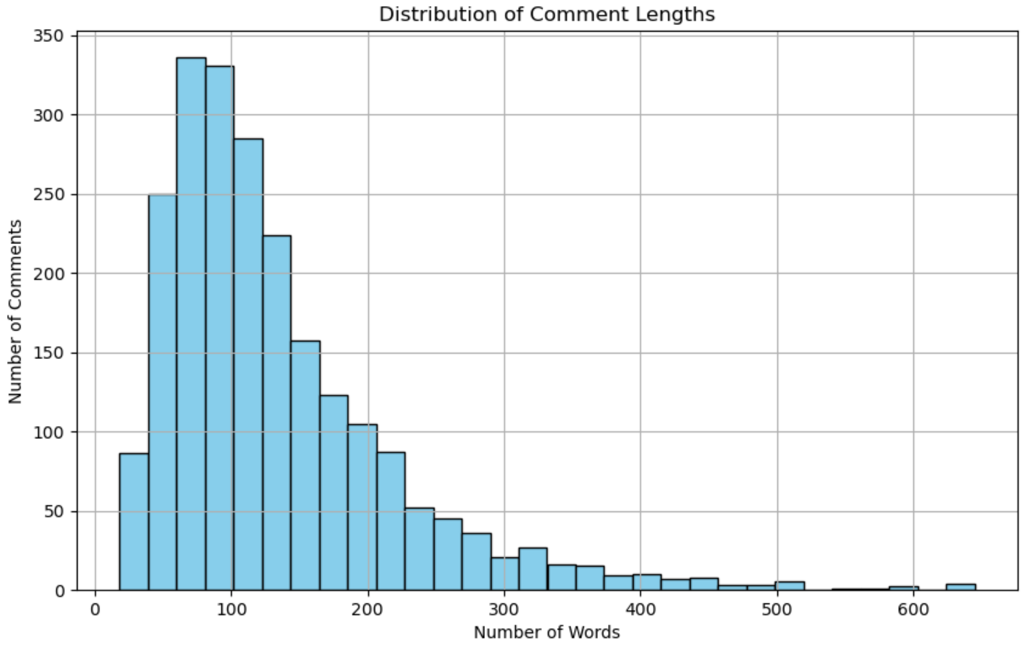
Exploring Ryanair Passenger Feedback: Topic Insights and Customer Satisfaction Classification

Università di Pisa

Niccolò Settimelli

Abstract

- **Increased Competition**: The European low-cost airline market is highly competitive, with Ryanair remaining a leader despite new significant competitors.
- **Damaged Corporate Image**: Recent developments have led to a rise in negative feedback and a tarnished reputation for Ryanair BBC News.
- **Stagnant Customer Satisfaction**: Customer satisfaction scores have remained stagnant over the past 4 years, showing no substantial improvement YouGov .



Ryanair Dataset

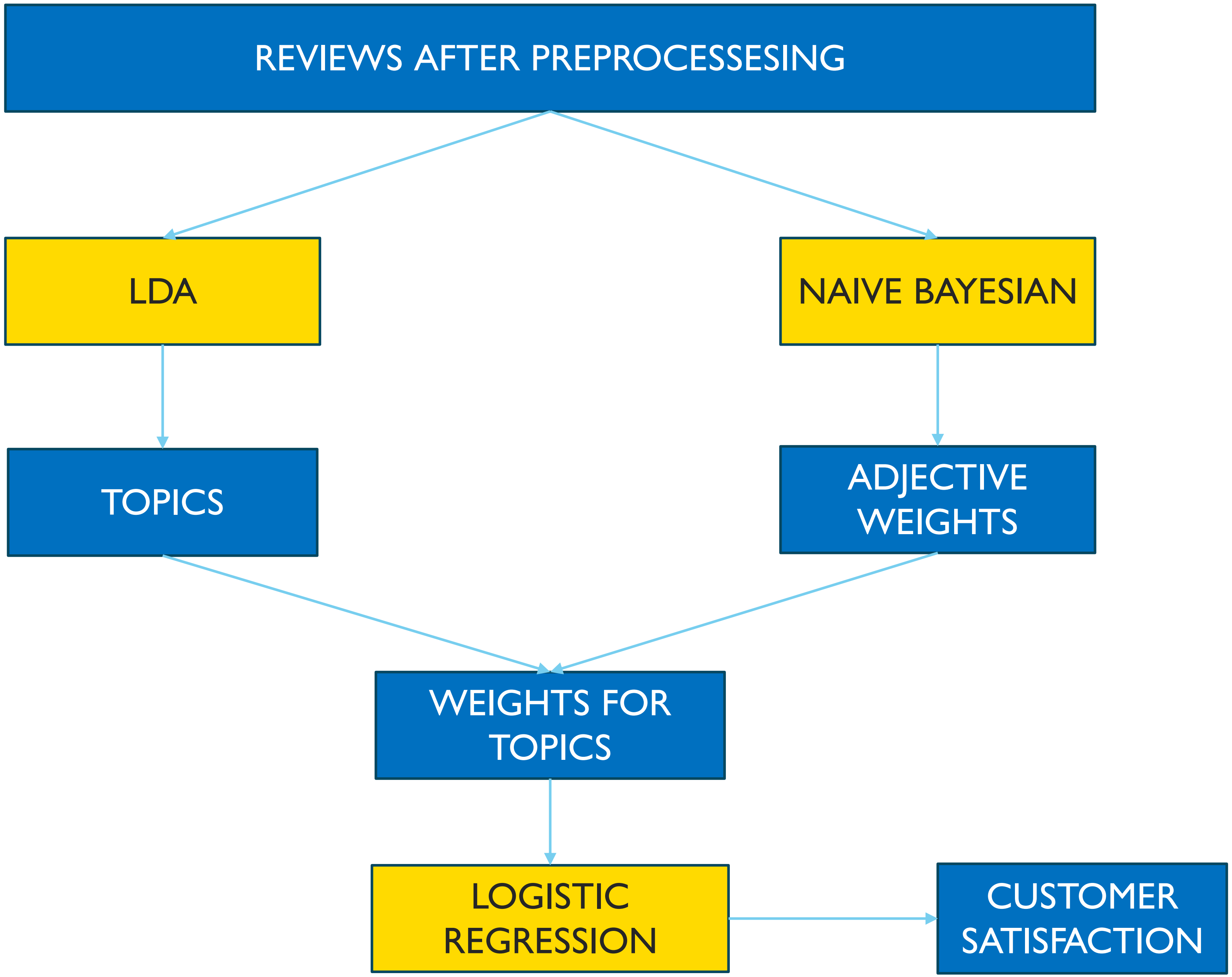
- Real reviews from **2012 to 2024**
- **2250** reviews
- **873** passengers will recommend the experience
- **1377** will not
- **130** average words for comment
- Not only reviews insight the dataset

Aim of the project

Topic Identification: Instead of traditional sentiment analysis, the project will identify and analyze key topics in Ryanair customer comments to predict sentiment and derive actionable insights.

Methodology: The approach includes preprocessing comments, using Latent Dirichlet Allocation (LDA) to find topics, evaluating these with Naive Bayes, and analyzing their impact on satisfaction with logistic regression [1].

Enhanced Insights: Pattern mining techniques will be applied to further explore topics, aiming to provide Ryanair with a detailed understanding of customer concerns and preferences for targeted improvements.


REVIEWS AFTER PREPROCESSESING
LDA
NAIVE BAYESIAN
TOPICS
ADJECTIVE WEIGHTS
WEIGHTS FOR TOPICS
LOGISTIC REGRESSION
CUSTOMER SATISFACTION

Text Mining preprocessing

Data Combination: Merge Comment Title and Comment fields and split into training and testing datasets.

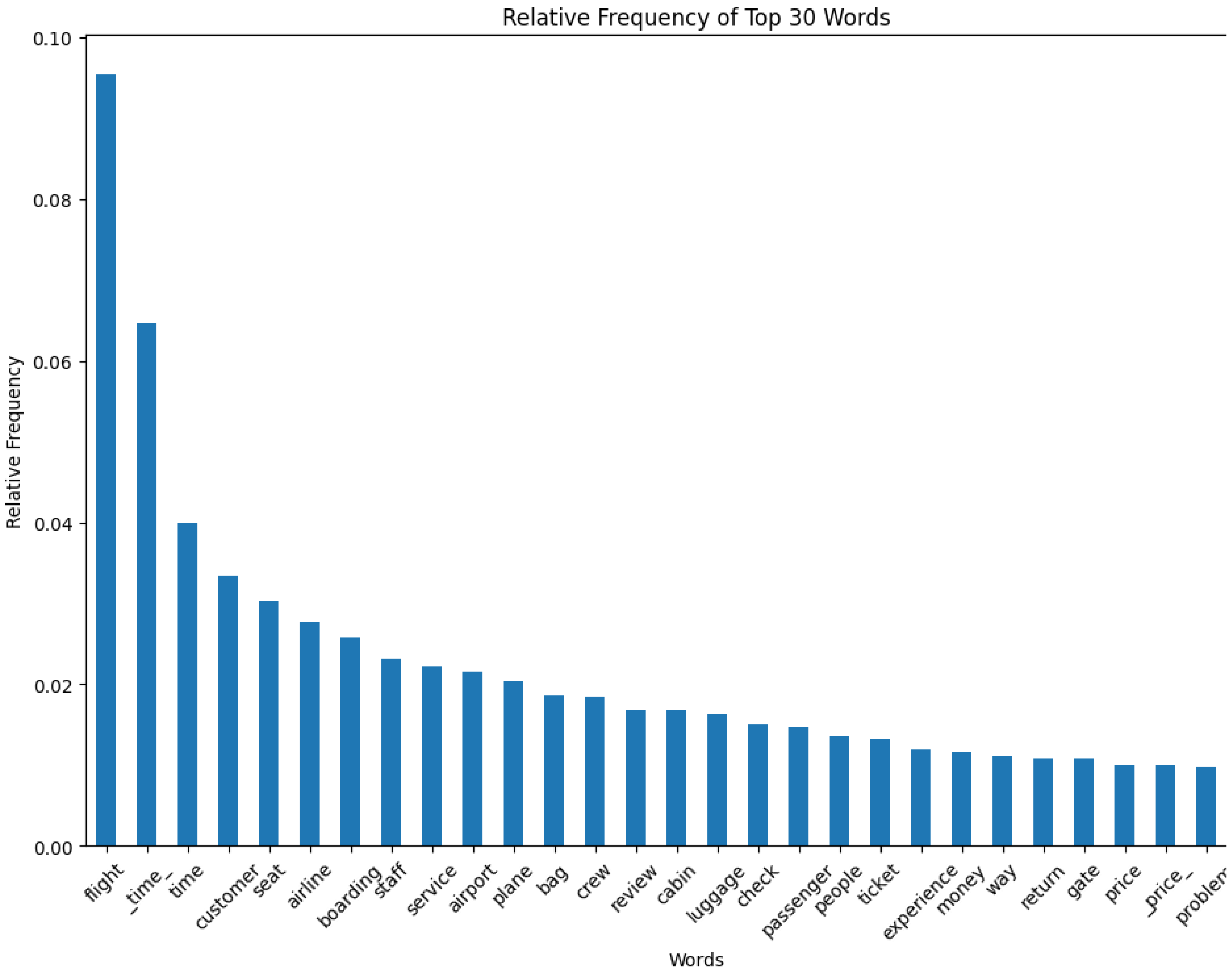
Text Processing: Apply CountVectorizer for tokenization, stop-word filtering, and stemming. Focus on nouns for LDA and adjectives for sentiment analysis.

BoW Construction: Build a Bag-of-Words (BoW) array with word frequencies, avoiding TF-IDF in favor of a simple frequency-based approach for LDA.

Replacement	Input Examples
night	21pm, 23:00, 10pm
morning	06am, 07:00, 08am
afternoon	01pm, 3:00pm, 5:00pm
evening	07pm, 8:00pm
time	30 min, 1 hour, days, hr
price	50, 100, 2000, 5000, USD
weight	10kg, 5 lbs, 100 grams, 2.
size	15 cm, 5 feet, 10 inches,
date	March 10, 2024, 10th Ma
city	Paris, Berlin, Rome
country	France, Germany, Italy
dayweek	Monday, Tue, Fri, Sun
airport	Heathrow, Charles de Ga
(removed)	1st, 2nd, 123

Word frequencies distribution

- Time combined with the key is the most frequent word
- Flight as we expected before is the most present word
- Also, the key _prize_ is a lot frequent



LDA phase

- Choose document length $N \sim Poisson(\xi)$
- Select topic distribution $\theta \sim Dir(\alpha)$
- For each word ω_n choose a topic $z_n \sim Multinomial(\theta)$ and generate ω_n from $\omega_n \sim p(\omega_n | z_n, \beta)$
- The final objective is to find the hidden parameters θ and z for maximizing this probability

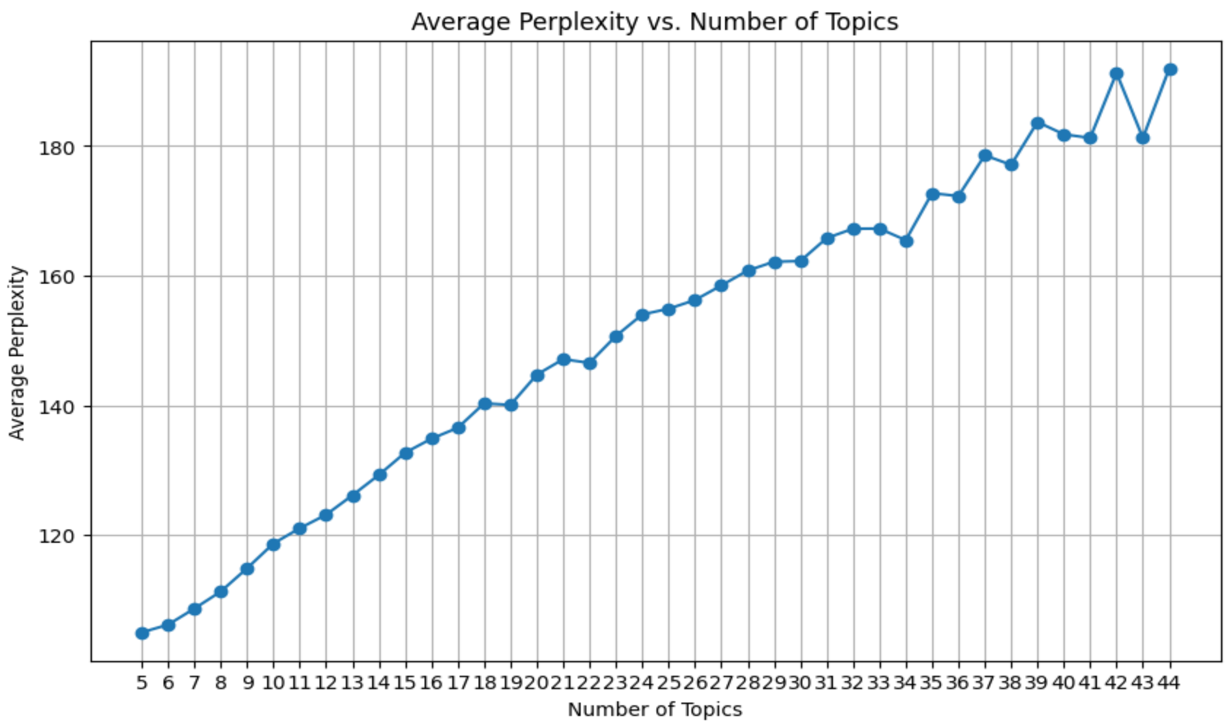
$$p(\theta, z | \omega, \alpha, \beta) = \frac{p(\theta, z, \omega | \alpha, \beta)}{p(\omega | \alpha, \beta)}$$

How to choice the number of topic

- **Perplexity Metric**: Measures model fit by evaluating word probabilities; helps determine the number of topics.

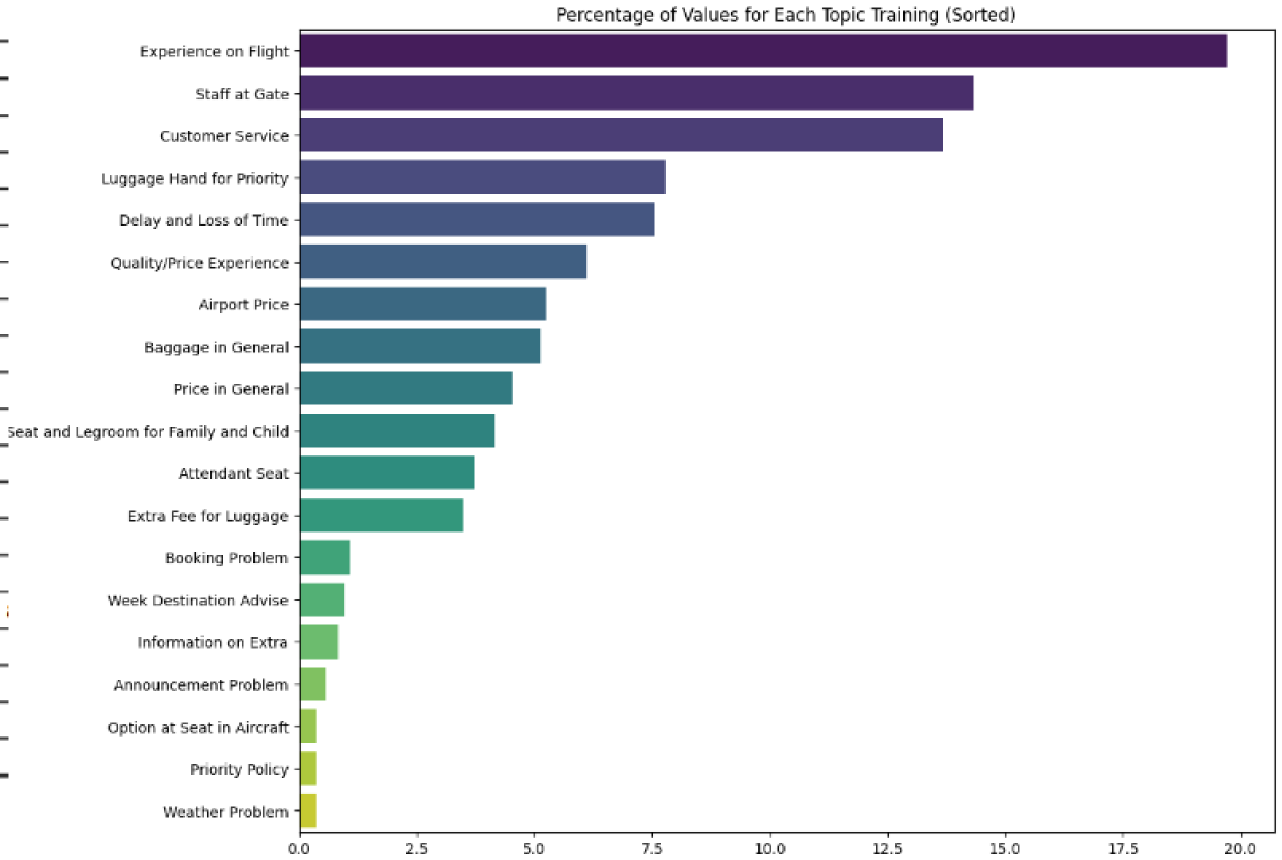
$$Perplexity = \exp(-\frac{\sum_{d=1}^{D} \log p(\omega_d|\theta_d,\phi)}{\sum_{d=1}^{D} N_d})$$

- **Challenges**: Limited dataset size made perplexity-based optimization impractical.
- **Chosen Topics**: Selected 19 topics based on the flattening of the perplexity curve, interpretability and threshold proposed in different research [2].



Topic Extraction and interpretation

Topic Number	Top Words	Interpretation
0	weather, kg, price, aircraft, employee	Weather Problem
1	_time_, flight, gate, staff, time	Staff at Gate
2	check, _price_, fee, agent, airport	Airport Price
3	flight, seat, aircraft, value, option	Option at Seat in Aircraft
4	flight, _time_, service, customer, time	Customer Service
5	seat, plane, front, row, attendant	Attendant Seat
6	luggage, hand, boarding, priority, flight	Luggage Hand for Priority
7	money, suitcase, value, water, flight	Extra Fee for Luggage
8	airline, experience, cost, budget, money	Quality/Price Experience
9	information, change, charge, drink, air	Information on Extra
10	bag, staff, cabin, baggage, flight	Baggage in General
11	booking, legroom, ground, fault, staff	Booking Problem
12	flight, crew, time, cabin, _time_	Experience on Flight
13	ticket, price, trip, luggage, time	Price in General
14	seat, leg, room, family, child	Seat and Legroom for Family
15	price, seat, problem, priority, policy	Priority Policy
16	week, destination, lot, case, time	Week Destination Advice
17	announcement, schedule, budget, person, airport	Announcement Problem
18	boarding, pass, customer, flight, _time_	Delay and Loss of Time



Adjective customer sentiment weight by Bayesian approach

Method: Use Naive Bayes to compute the conditional probability of a class (e.g., recommendation) given an adjective.

$$P(Class|Adjective) = \frac{P(Adjective|Class) \cdot P(Class)}{P(Adjective)}$$

Purpose: Determine how the presence of specific adjectives affects the likelihood of a recommendation in reviews.

Adjustment: Normalize the computed weights to a range of -1 to 1 for consistent interpretation.

$$Normalized\ Weight = \frac{X - \min(X)}{\max(X) - \min(X)} \cdot 2 - 1$$

```
Top 10 Positive Adjectives:
    Adjective    Weight
50       good  1.000000
47   friendly  0.793954
16      cheap  0.664202
51      great  0.616184
31  efficient  0.610123
92   pleasant  0.578311
44      first  0.571624
54    helpful  0.536337
71        low  0.497516
38      extra  0.489263

Top 10 Negative Adjectives:
          Adjective    Weight
26            dirty -0.644585
135  unprofessional -0.731139
```

Assessement topic with adjective scores

- **Adjective Normalization**: Focus on adjectives with high weights (above 0.75 or below -0.75) to capture significant sentiment influences.

- **Sentiment Score Formula**:

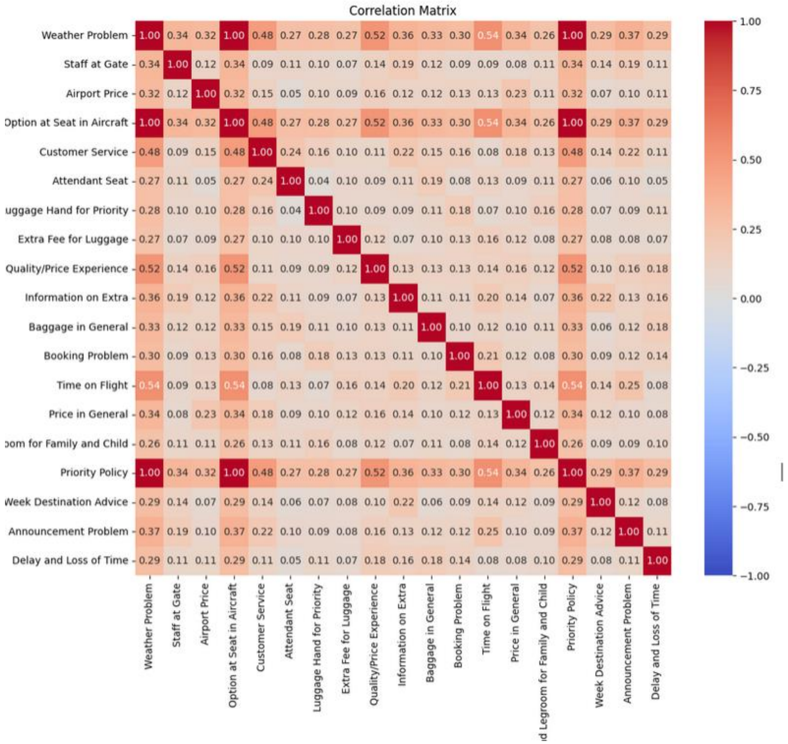
$$Sentiment\ Score = Adj\ Frequency\ x\ Weight\ of\ Adj$$

- **Topic Redistribution**: Sentiment scores are allocated to topics based on their probability in the review.

Classification with logistic regression

- **Logistic Regression:** Chosen for its effectiveness with numerical data and interpretability of feature impacts

- For **feature selection** we observe before Correlation matrix, but all the topics are less or more independent and than we try Sequential Feature Selection



Logistic regression results

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.1826	0.068	-17.325	0.000	-1.316	-1.049
Staff at Gate	1.2080	0.341	3.539	0.000	0.539	1.877
Airport Price	-1.3474	0.798	-1.689	0.091	-2.911	0.216
Customer Service	2.1149	0.396	5.338	0.000	1.338	2.891
Luggage Hand for Priority	3.3816	0.616	5.490	0.000	2.174	4.589
Quality/Price Experience	2.3033	0.682	3.378	0.001	0.967	3.640
Time on Flight	4.9388	0.359	13.758	0.000	4.235	5.642
Seat and Legroom for Family and Child	3.9888	1.034	3.856	0.000	1.961	6.016
Week Destination Advice	2.1945	2.607	0.842	0.400	-2.915	7.304

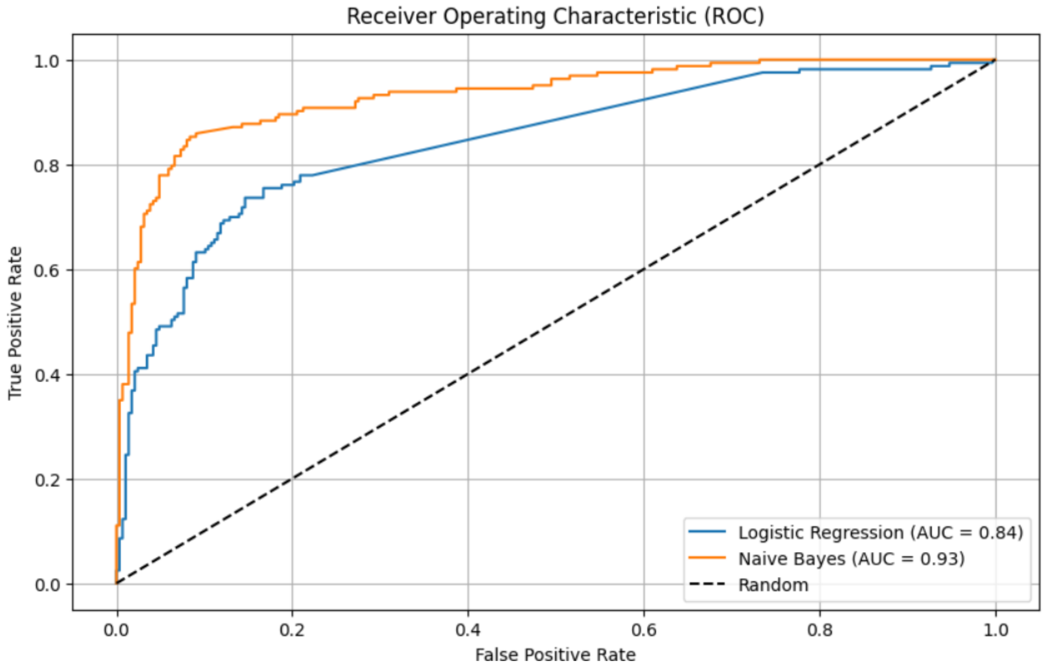
Classification Report:

	precision	recall	f1-score	support
0	0.80	0.92	0.86	287
1	0.80	0.60	0.69	163
accuracy			0.80	450
macro avg	0.80	0.76	0.77	450

Comparison with a classical Naive Bayesian classifier

- **Model Performance:** Logistic Regression using topic shows a general decline in performance metrics compared to Naive Bayes, but the reduction is moderate enough to be outweighed by the improved interpretability of the model.
- **ROC Curves:** ROC curves reveal performance differences between the two models, highlighting trade-offs between accuracy and model interpretability.

```
ROC AUC Logistic Regression: 0.84
ROC AUC Naive Bayes: 0.93
```

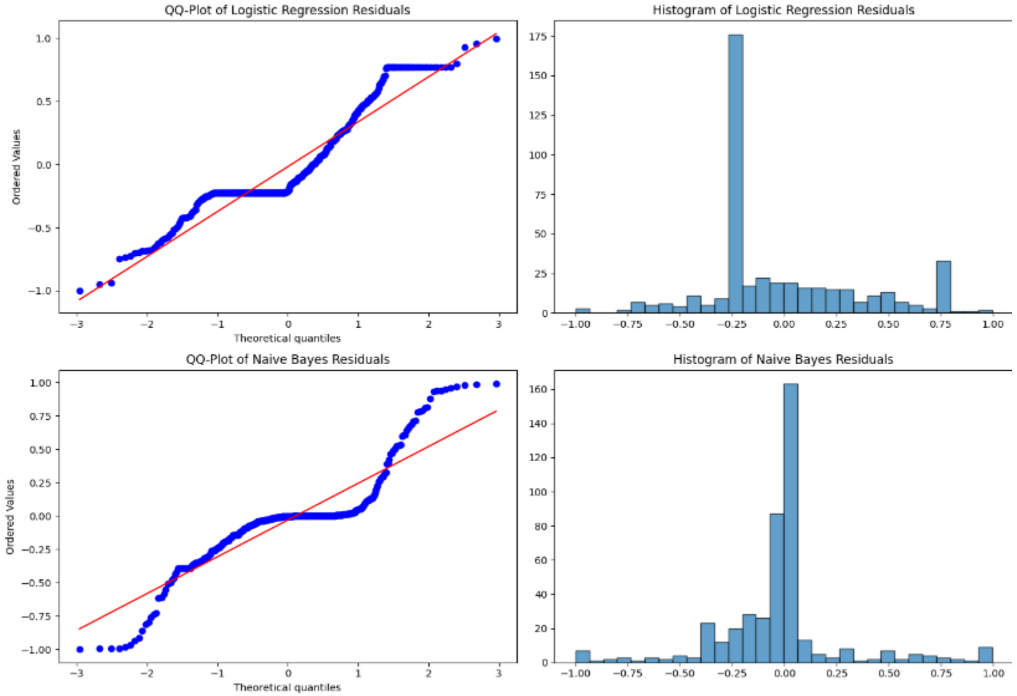



```
Logistic Regression Classification Report:
              precision    recall  f1-score   support

           0       0.79      0.92      0.85       287
           1       0.80      0.58      0.67       163

    accuracy                           0.80       450
   macro avg       0.80      0.75      0.76       450
weighted avg       0.80      0.80      0.79       450

Naive Bayes Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.92      0.91       287
           1       0.86      0.83      0.84       163

    accuracy                           0.89       450
   macro avg       0.88      0.88      0.88       450
weighted avg       0.89      0.89      0.89       450

ROC AUC Logistic Regression: 0.84
ROC AUC Naive Bayes: 0.93
```

Cross-validation and statistical residual analysis

- **Cross-Validation:** K-fold cross-validation was used to assess the reduction in accuracy, offering a detailed view of model performance and consistency across various data subsets.
- **Residual Analysis:** Comparison of residuals showed a non-Gaussian distribution, with the Wilcoxon test confirming a performance decline when moving from Naive Bayes to Logistic Regression.
- **Performance Trade-Off:** Despite a reduction in accuracy, the Logistic Regression model's interpretability benefits are a notable advantage over the Naive Bayes model.

```
Logistic Regression Cross-Validation Accuracy: 0.768 ± 0.008
Naive Bayes Cross-Validation Accuracy: 0.840 ± 0.014
Optimization terminated successfully.
         Current function value: 0.493847
         Iterations 7
```




```
Wilcoxon test statistic: 49288.000
Wilcoxon test p-value: 0.599
```

Graph Interface

Enter your comment:

I had a terrible experience with Ryanair Airlines.
The flight was significantly delayed with no clear
explanation, and the staff didn't seem
particularly interested in resolving the issue.
Additionally, the airplane was dirty and the seat

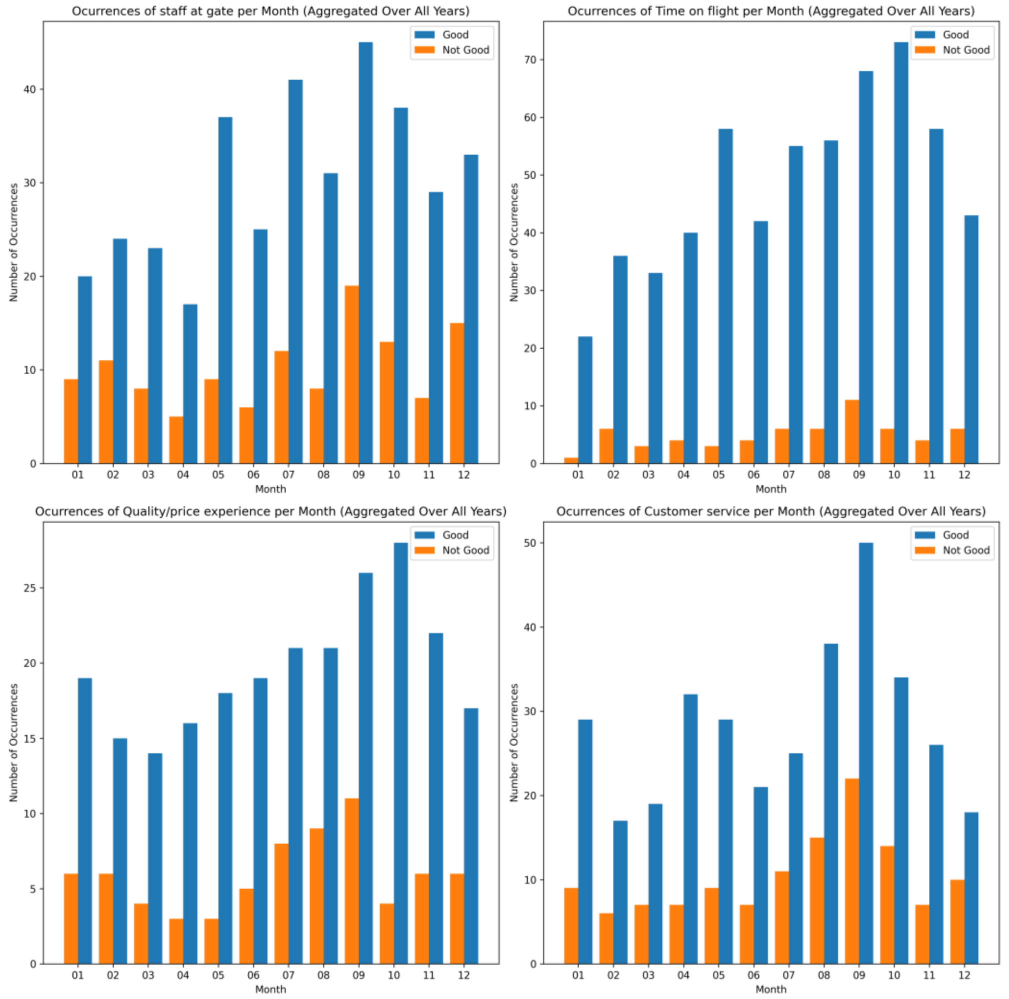
Analyze

Recommendation: No

Topic	Coefficient	P-Value
Quality/Price Experience	-4.65052595770986	0.000851560679482217
Seat and Legroom for Family and Ch	-4.562410904154383	0.005321649582708577
Priority Policy	3.389766176754014	0.018572048684664878
Weather Problem	-3.6616931441471365	0.022253850165984614
Information on Extra	-4.105092175948172	0.03303826051602398
Extra Fee for Luggage	2.553427583737859	0.037581205128974775
Attendant Seat	-0.5101025611537944	0.05296655316508915
Luggage Hand for Priority	-3.929088072071909	0.055421457192701
Booking Problem	4.845199681210147	0.05804381161805571
Delay and Loss of Time	-1.3193278625927451	0.06272653462231062

Preprocessing on new ryanair_customer_satisfaction dataset

- **Dataset Optimization**
  Removed irrelevant features and those with excessive missing values, such as "Wifi Entertainment” or “Aircraft”
- **Topic Discretization**
  Categorized topic scores into "Good," "Not Good," and "Not Interesting" to manage sparsity in the data.
- **Statistical Visualization**
  Analyzed monthly frequencies of "Good" and "Not Good" ratings for key topics to identify seasonal trends in customer satisfaction.



Using our topic for different applications

- **Utilization of Identified Topics**
  Leverage identified topics with customer travel information from reviews.

- **Frequent Pattern Mining**
  Apply frequent pattern mining to associate topics with travel characteristics using association rules.

- **Infinite Applications for Ryanair**
  Potential uses include personalized marketing, targeted service improvements, and enhanced customer insights.

Pattern mining example

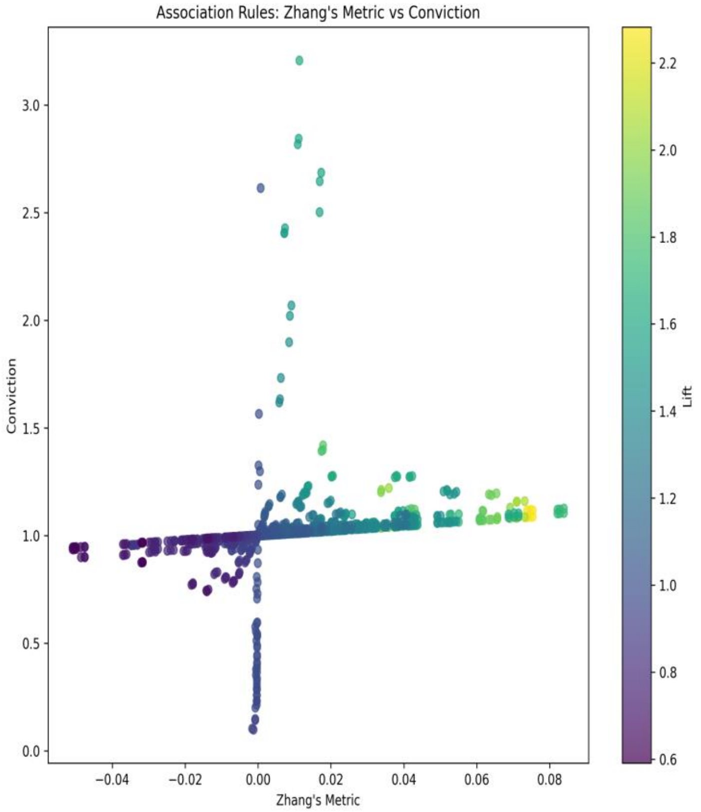
- **Feature Integration and Pattern Mining**
  Integrated the following features for pattern mining: **Type of Traveller**, **Origin, Destination, Passenger Country, Seat Type**, **Month**
- Applied the Apriori algorithm to discover frequent patterns, focusing on 'Good' and 'Not Good' labels while excluding 'Not Interesting' labels.
- **Pattern Validation Metrics:** We need more complex metrics because our interesting items are rare
  - **Lift**: Measures how much more likely Y is given X compared to its baseline probability.
  - **Zhang's Metric**: Assesses the strength of association between X and Y relative to X's baseline probability.
  - **Conviction**: Evaluates the degree of dependency between X and Y, with values greater than 1 indicating a strong association.

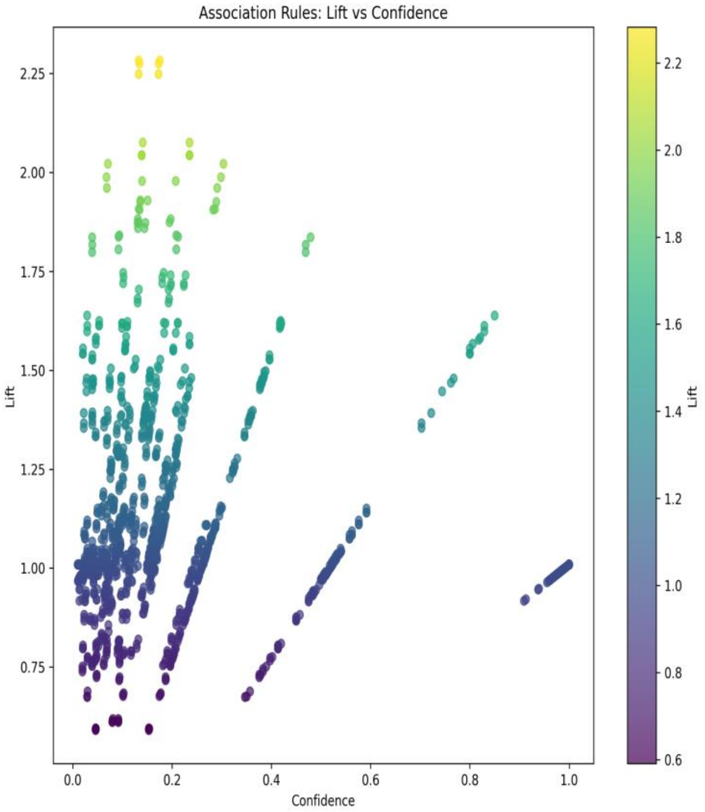
Patter mining results

Metric Analysis
Plotted metrics to identify patterns with the best trade-off between confidence and complex metrics, targeting confidence > 0.75 and lift > 1.6.

Examples of Rules Above Threshold

- **Rule 1**: Travelers from Stansted Airport with 'Good' quality-price ratings have high confidence and lift.
- **Rule 2**: Positive feedback in Economy Class often correlates with other factors rather than service quality, showing strong lift and confidence metrics.





Conclusion

- **Effective Topic Analysis**
  - Employed LDA and logistic regression to achieve 77% accuracy in sentiment prediction, but finding different customer satisfaction dimensions.
- **Actionable Insights**
  - Uncovered key trends and patterns as example of our principal object, such as improved quality-price perception at certain airports.
- **Inspired by Research**
  - Adapted from [1] with modifications for dataset specifics and future comparisons.

References

[1] Lucini, F. R., Tonetto, L. M., Fogliatto, F. S., & Anzanello, M. J.

- **Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews**
- *Journal of Air Transport Management*, 2020, 83: 101760.
- https://doi.org/10.1016/j.jairtraman.2019.101760

[2] Bian, J., Yoshigoe, K., Hicks, A., Yuan, J., He, Z., Xie, M., ... & Modave, F. (2016). **Mining Twitter to assess the public perception of the “Internet of Things”**. *PloS one*, *11*(7), e0158450.

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0158450

Dateset link:

- https://www.kaggle.com/datasets/cristaliss/ryanair-reviews-ratings

Thank You for the attention

Niccolò Settimelli