

Ideatory MakeMyTrip Submission

Submitted by

Name: Avinash Singh Pundhir

Ideatory username:
meavinash100@gmail.com

Best Airlines

- Which are the best 2 airlines?
 - Asiana Airlines
 - Eva air

ASIANA AIRLINES 

EVA  AIR

What passengers liked

- What did passengers like the most about the best two airlines?
 - Comfortable seats
 - Quality of service provided by crew



Worst Airlines

- Which are the worst 2 airlines?
 - United Airlines
 - Vueling Airlines



What passengers disliked

- What did passengers dislike the most about the worst two airlines?
 - Flights not on time
 - Quality of seats



Approach

- The following are the parameters used to identify top and worst airlines:
 - Overall rating
 - Weighted rating (derived based on text reviews)
 - % of Recommended to total reviews
 - Rate of change in rating over time
- Weights assigned to various ratings identified using industry market research and text reviews
- Text reviews are used to identify weights assigned to various ratings
- Text reviews are also used to identify what passengers like and dislike most about best and worst airlines

Approach

- The following are the derived weights for various ratings:
 - Seat comfort: 0.3
 - Cabin Staff: 0.2
 - Food and beverages: 0.2
 - Entertainment: 0.1
 - Grounds services: 0.1
 - WiFi Service: 0.0
 - Value for money: 0.1



These weights are associated based on tourism industry studies and text mining of reviews.

How did you do it?

- Problem definition:
 - Identify two best and worst airlines based on the provided data
 - Data contains overall rating, ratings across various parameters, recommended values and user reviews
- Data preparation:
 - Missing values are replaced by the average values across the parameters for the specific airlines
 - % Recommended to total review column is added across various airlines
 - Calculation of total number of data rows available across different airlines to facilitate final rating calculations

How did you do it?

- Overview of implementation
 - Final rating is calculated using the following approach:
 - $\text{Final rating} = 0.5 * \text{over all rating} + 0.5 * \text{weighted rating}$
 - Recommended %
 - Ranks are identified based on consolidated performance across final ratings and recommended %
- Weighted ratings are calculated by assigning different weights to different ratings
- These weights are identified by text mining user reviews for different features
- Finally to identify top 2 and bottom 2 analysis time series analysis is used to compare airline performance across review period
- To identify most liked and most disliked features we have mined the reviews for top 2 and bottom 2 airlines for frequency of words

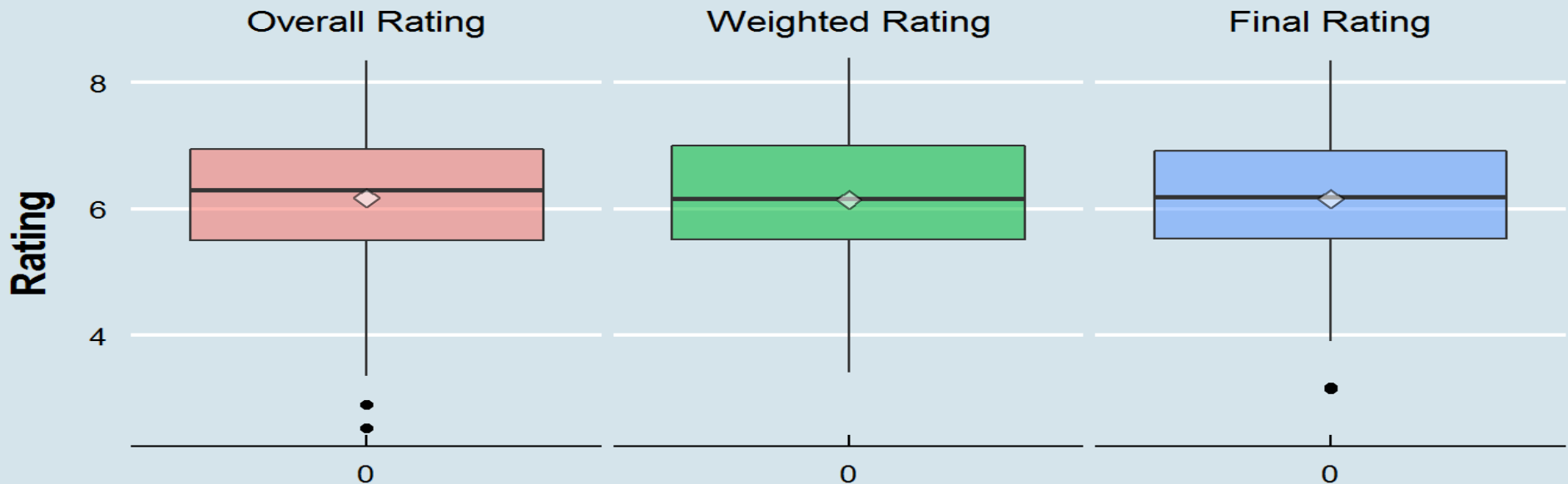
How did you do it?

- Visualization of results

Exploratory Analysis of Calculated Ratings

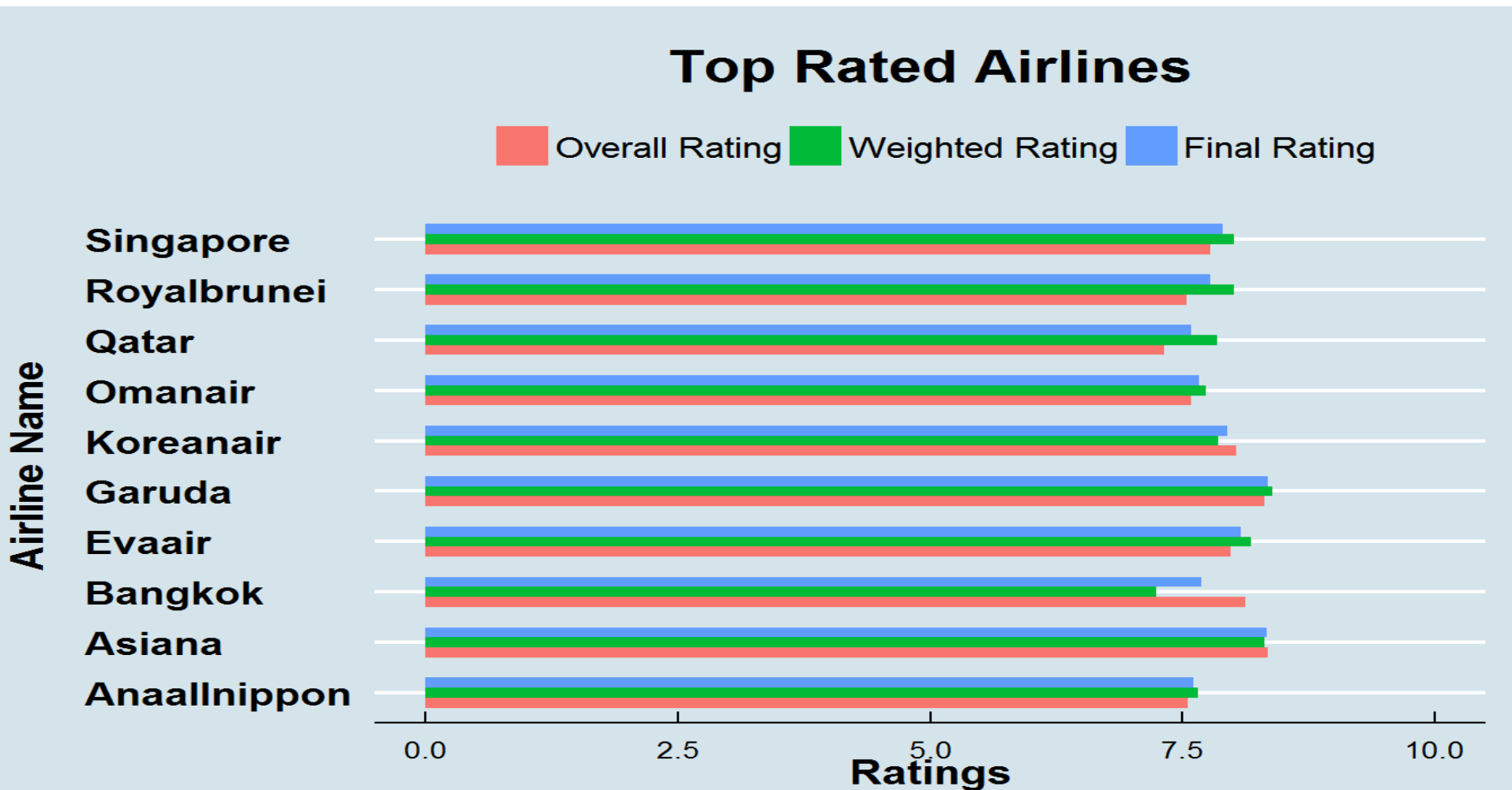
Overall Rating Weighted Rating Final Rating

0.9



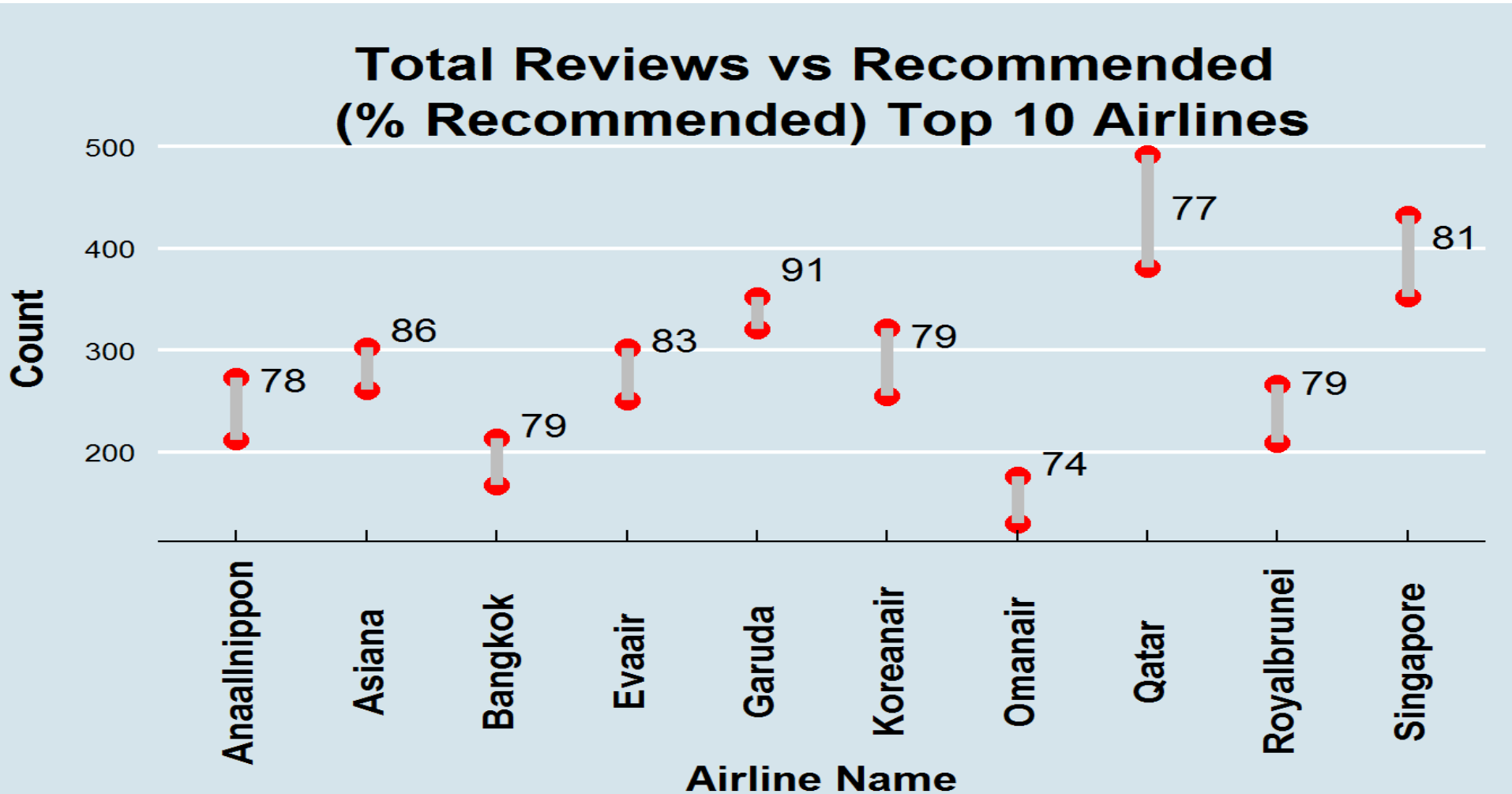
How did you do it?

- Visualization of results



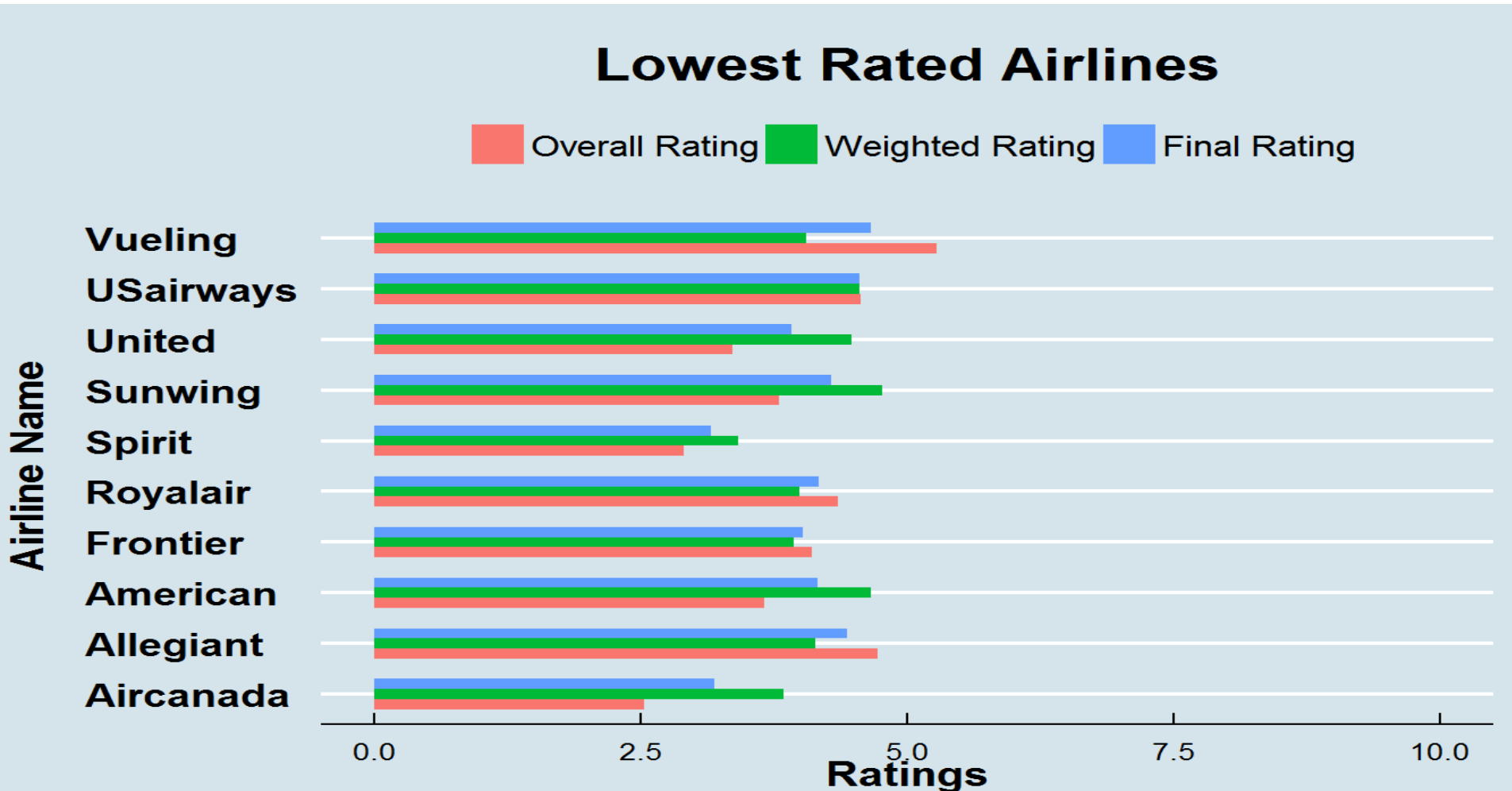
How did you do it?

- Visualization of results



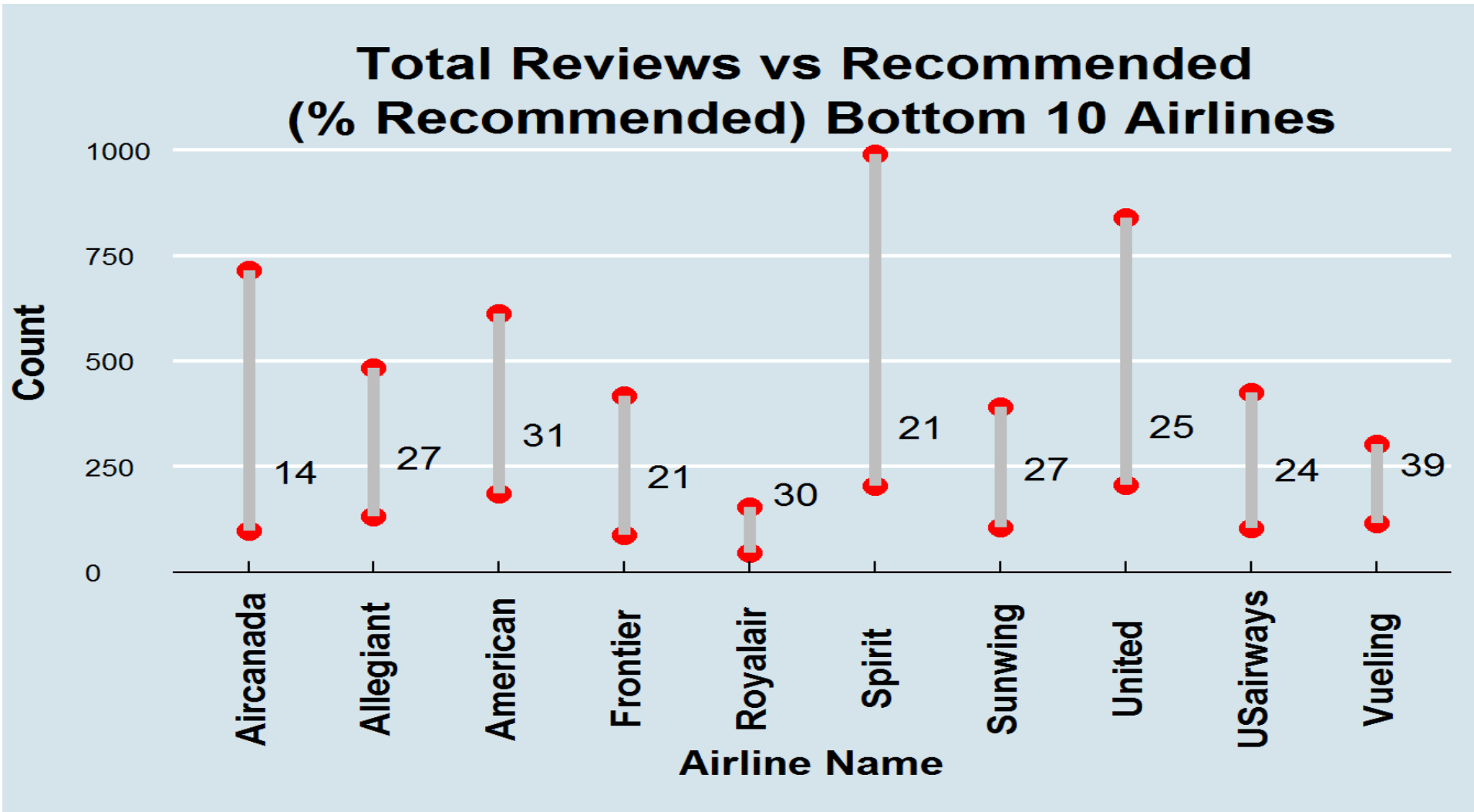
How did you do it?

- Visualization of results



How did you do it?

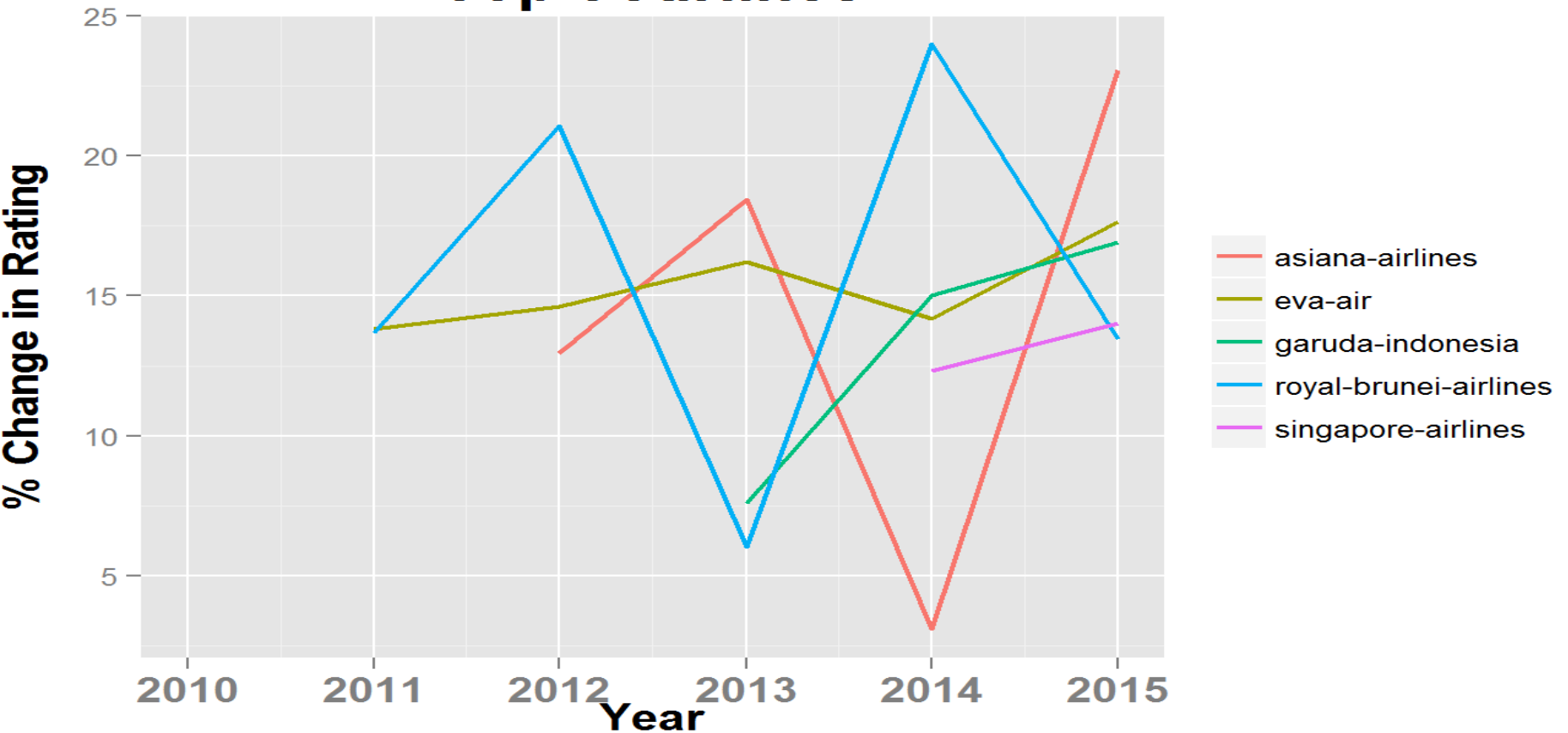
- Visualization of results



How did you do it?

- Visualization of results

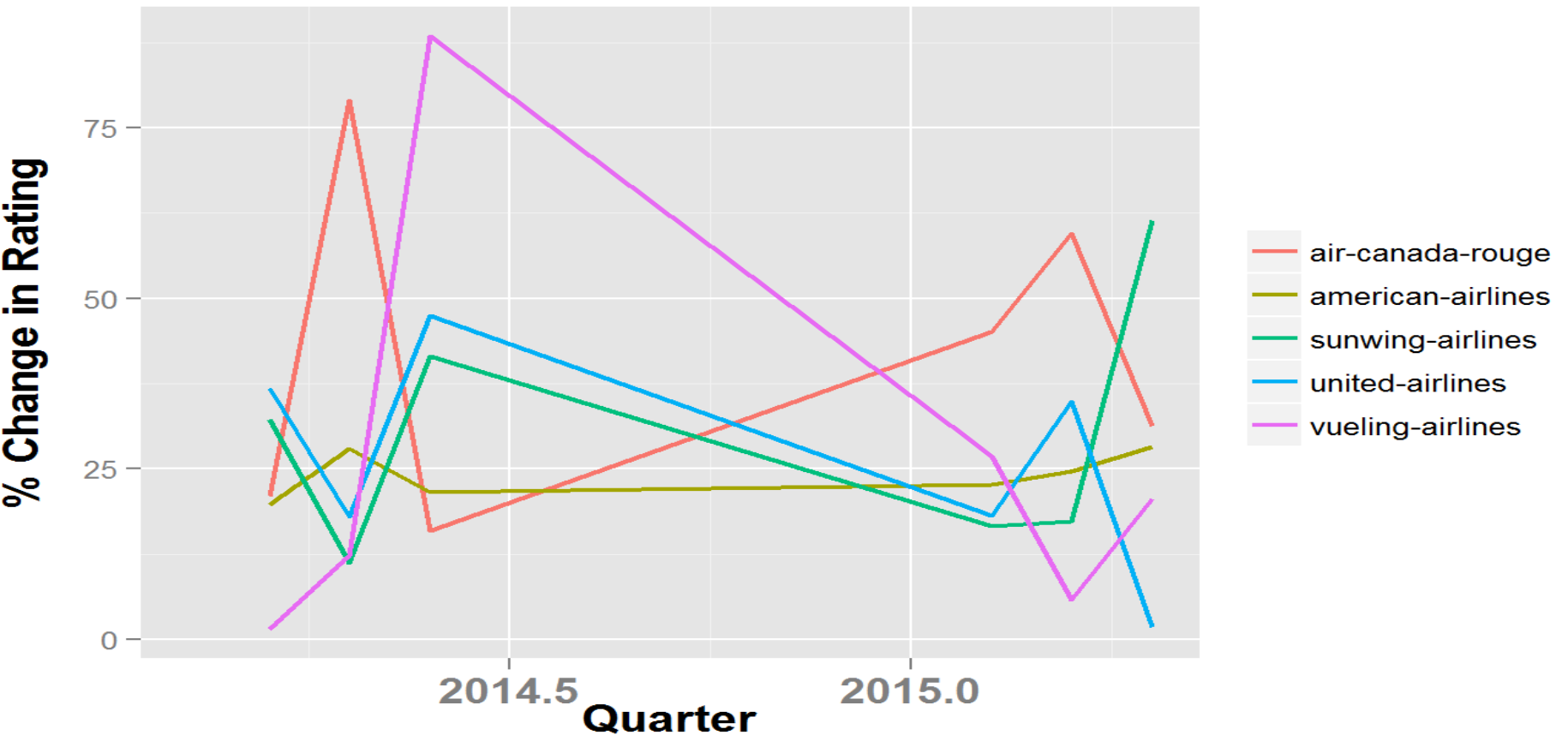
**Change in Rating Over Time
Top 5 Airlines**



How did you do it?

- Visualization of results

**Change in Rating Over Time
Bottom 5 Airlines**



How did you do it?

- Issues encountered:
 - Normalization of data to compare as we have varying number of reviews across different airlines
 - Imputing of missing values
 - Identification of minimum number of data records to include any airline in analysis
- Other approaches tried
 - Bayesian sentiment analysis to categorize user text reviews in categories like positive, negative and neutral

Tools

- What tools/software did you use?
 - R Programming
 - R Studio
 - Various packages in R to facilitate analysis