Tweets Data: Insight Report and Analysis

By Darron Kotoyan

Appendix

Data Assessment

- Not many numerical values, seems like the most useful data is written
- Data is messy, missing values
- Data is not consistent: Two people can be tweeting from Los Angeles but will be written in multiple different ways
- Written information is unique and not easy to work with
- **Delta's tweets are all "@JetBlue", therefore there is an error and "Delta" should be switched to "Jet Blue"**

Delta		SL_024		0	@JetBlue 653. Just fo
Delta		SL_024		0	@JetBlue no I asked.
Delta		_PauLiNHa9		0	@JetBlue 2324 from C
Delta		Tay__Mc		0	@JetBlue Yes, I filed y
Delta		SL_024		0	@JetBlue shocker my
Delta		ManuelNT		0	@jetblue as courtesy t
Delta		rachelicha		0	@JetBlue Are you havi

Steps Taken

Data Cleaning:
- Changed Delta name to Jet Blue
- Removed negative_reason_gold and airline_sentiment_gold
- Dropped NA values

Steps Taken

General Analysis:

- Looked at count of each negative reason compared to every airline
- Looked at % of each negative reason compared to every airline
- Looked at count of sentiment per airline
- Found airline sentiment percentages

Steps Taken

Sentiment Analysis

- Compared average airline sentiment confidence and average negative reason confidence per airline
- Time series analysis
- Analyzed sentiment of tweet with time spikes

General Analysis

- Looked at the count of negative reviews in comparison to the airline
- **United had the most negative reviews**

Count of airline_sentiment Row Labels	American	Jet Blue	Southwest	United	US Airways	Virgin America	Grand Total
(blank)	799	1267	1234	1189	650	323	5462
Customer Service Issue	768	199	391	681	811	60	2910
Late Flight	249	269	152	525	453	17	1665
Can't Tell	198	186	159	379	246	22	1190
Lost Luggage	149	57	90	269	154	5	724
Bad Flight	87	64	90	216	104	19	580
Cancelled Flight	246	51	162	181	189	18	847
Flight Attendant Complaints	87	60	38	168	123	5	481
Flight Booking Problems	130	44	61	144	122	28	529
longlines	34	14	29	48	50	3	178
Damaged Luggage	12	11	14	22	11	4	74
Grand Total	**2759**	**2222**	**2420**	**3822**	**2913**	**504**	**14640**

- Compared negative complaints to each airline
- Jet Blue leads in blank complaints
- United leads in Late flights, lost luggage, bad flights, flight attendant complaints, and damaged luggage
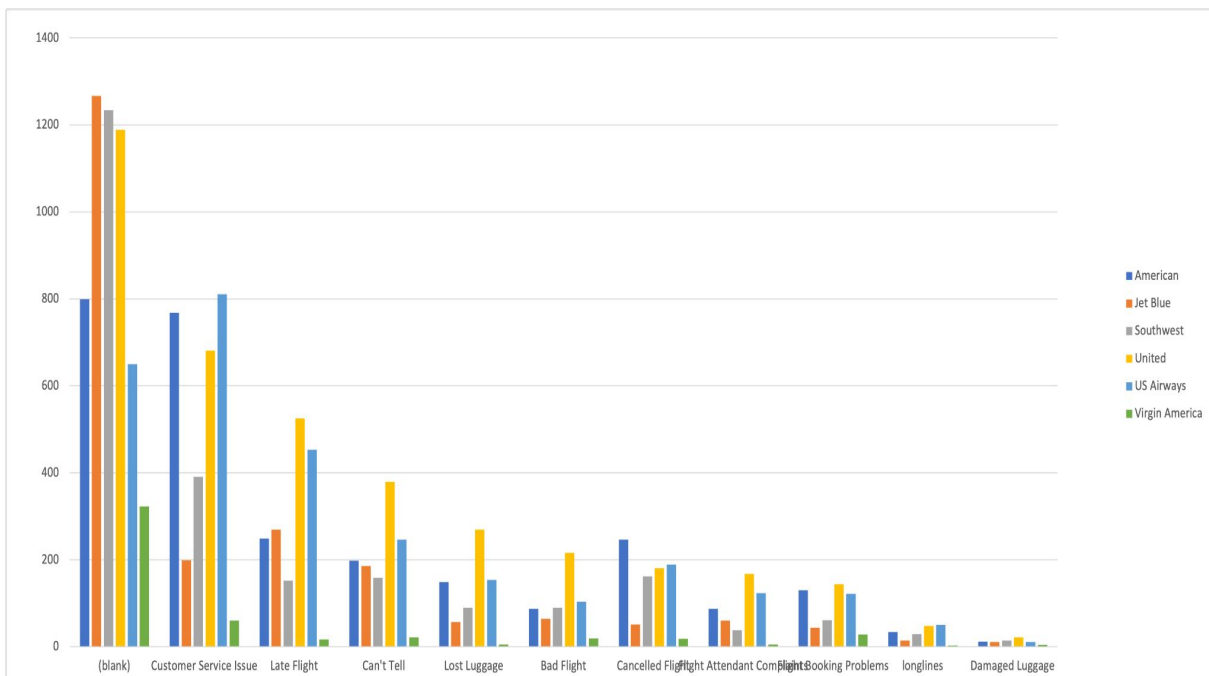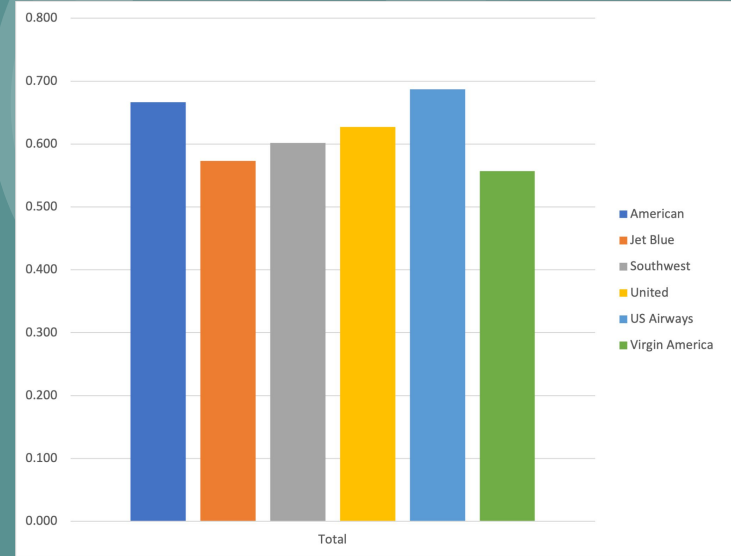
Sentiment confidence analysis

- Average sentiment confidence score for each airline
- U.S Airways has the **highest** average sentiment confidence score at .922
- American has the **second highest** average sentiment confidence score at .917
- At **third highest**, United has an average sentiment confidence score of .901
- Others closely follow
- **Meaning that generally, customers are highly confident in the accuracy of the sentiment scores that they have given about these airlines**

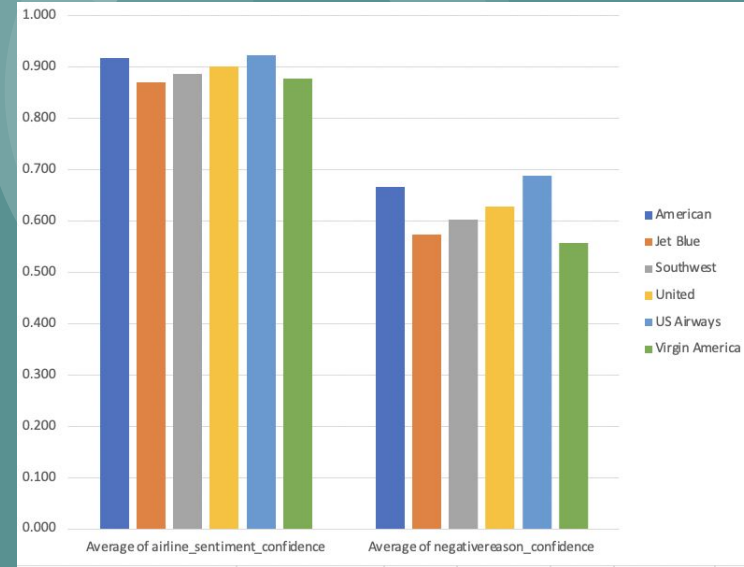| Values | Column Labels ▼ | | | | | | |
| | American | Jet Blue | Southwest | United | US Airways | Virgin America | Grand Total |
| Average of airline_sentiment_confidence | 0.917 | 0.870 | 0.887 | 0.901 | 0.922 | 0.876 | 0.900 |
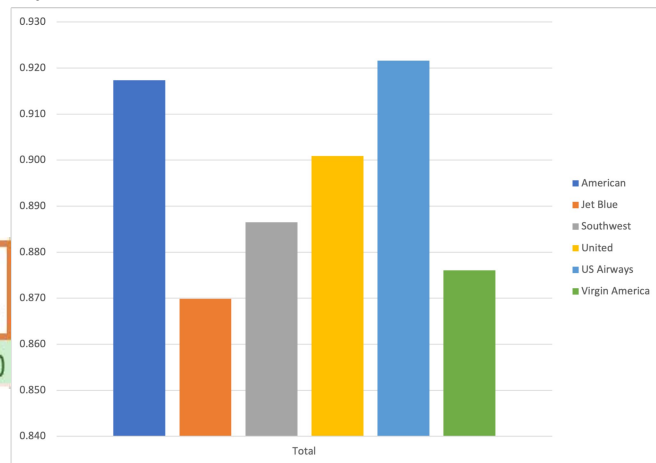
Let's look at negative reason confidence

- Generally, customers are less confident in the accuracy of their negative reasons for an airline
- US Airways, American, and United customers have the most confidence in their negative reasons for the airline

	Column Labels						
	American	Jet Blue	Southwest	United	US Airways	Virgin America	Grand Total
Average of negativereason_confidence	0.666	0.573	0.602	0.627	0.687	0.557	0.638

Negative reason confidence vs Sentiment confidence

- We can clearly see that the average airline sentiment confidence is significantly higher than negative reason confidence!
- What does this mean?
- It tells us that generally, customers significantly believe more in the validity of their sentiment rather than their negative reason
- Focusing on sentiment confidence will give more accurate results of how the customer feels!

Values	Column Labels						
	American	Jet Blue	Southwest	United	US Airways	Virgin America	Grand Total
Average of airline_sentiment_confidence	0.917	0.870	0.887	0.901	0.922	0.876	0.900
Average of negativereason_confidence	0.666	0.573	0.602	0.627	0.687	0.557	0.638

Going Deeper into sentiment confidence

- *US Airways has the highest percentage of negative sentiment and lowest percentage of positive sentiment out of all airlines*
- *American has second highest percentage of negative sentiments*
- *United has third highest percentage of negative sentiments*

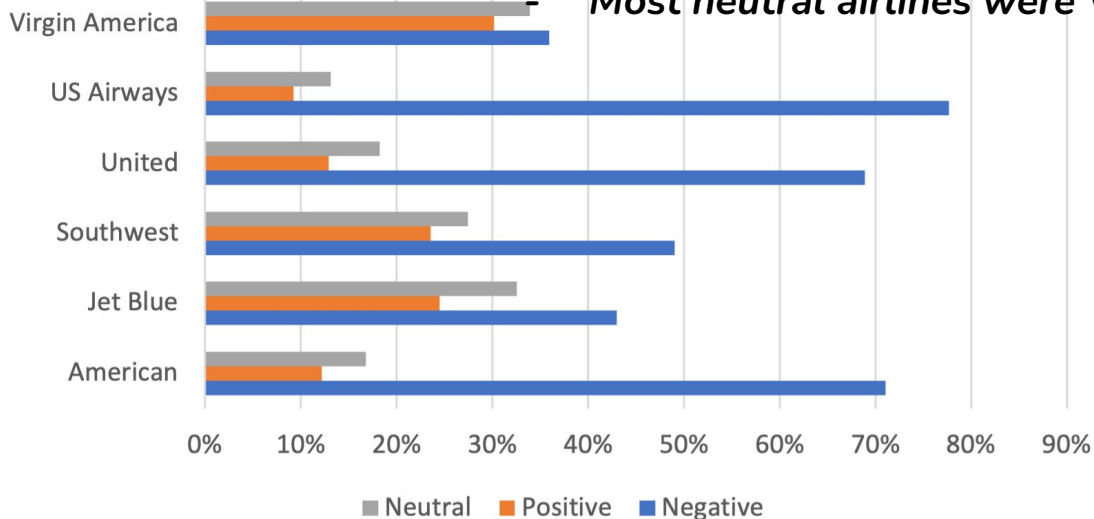|	American	Jet Blue	Southwest	United	US Airways	Virgin America
Negative	71%	43%	49%	69%	78%	36%
Positive	12%	24%	24%	13%	9%	30%
Neutral	17%	33%	27%	18%	13%	34%

Going Deeper into sentiment confidence

- *Virgin America has highest percentage of positive sentiments and lowest percentage negative sentiments*
- *American and Southwest are tied for second highest percentage of positive sentiments*
- *Most neutral airlines were Virgin America, Jet Blue, and Southwest*

Sentiment Per Airline

Chart: Sentiment Per Airline — horizontal bar chart showing Neutral, Positive, and Negative percentages for Virgin America, US Airways, United, Southwest, Jet Blue, and American, with an x-axis from 0% to 90%.

Legend: ■ Neutral ■ Positive ■ Negative

Looking at Time Series

- Comparing tweet volume and date of creation
- See a high spike in tweet volume for all airlines on Feb 22-23
- Factored in sentiment, and we see extremely high spike in negative sentiment on these dates as well
- Leads me to believe something very negative affected airlines in this period of time that has caused such outrage
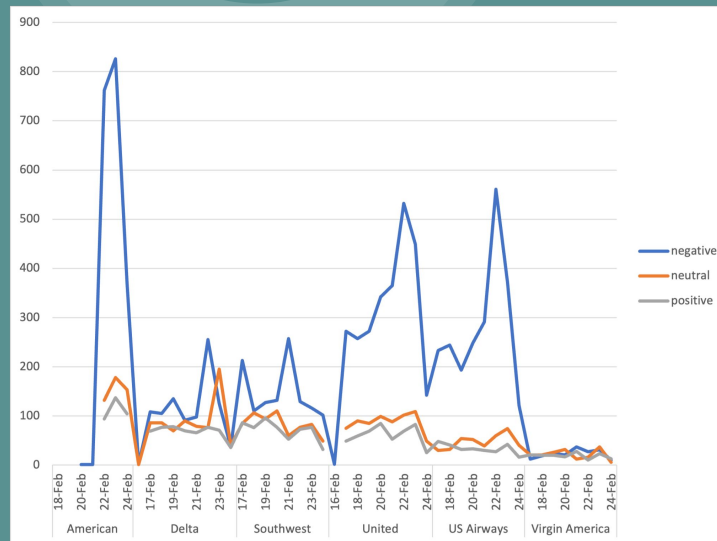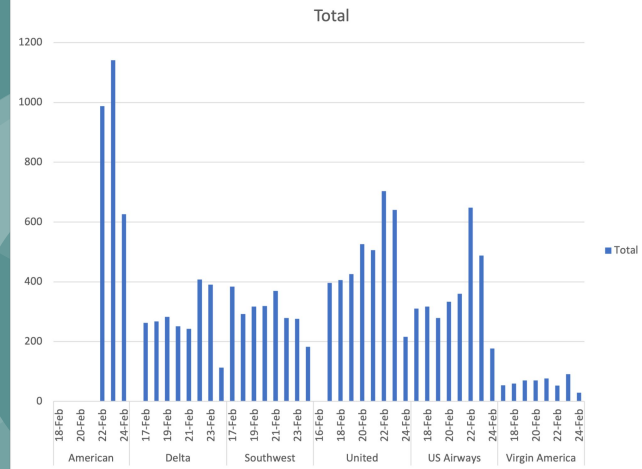
Executive Summary

- In the period of time given, the worst performing airlines sentiment wise, were US Airways, American, and United
- The best performing airlines sentiment wise, were Virgin America, American, and Southwest
- There was a spike of negative sentiment on Feb 22-23, leads me to believe it's due to outside factors
- Late flights, lost luggage, bad flights, flight attendant complaints, and damaged luggage are leading complaints, fix these and make more money with happier customers