Aurora Borealis

Katie Briggs

Bellevue University

Abstract

The Aurora Borealis is a phenomenon observed in the northern on our planet and is the result of collisions between gas particles in the Earth's atmosphere [1]. They are sometimes referred as the Northern Lights and described as bright dancing ribbons of light.   The southern hemisphere calls the phenomenon Aurora Austrailis. This analysis will be based on the northern area, specifically in the Alaska location.   The variations of color are a result of oxygen molecules in various locations above the earth. Sunspot activity occurs and the solar winds blow electrons and protons toward the earth and these particles are charged by the earth's magnetic field. Space weather forecasts include the aurora and are measured by their kp index which ranges from 0 to 9, with 1 being calm and 6 being highly likely to view the Northern Lights. Planetary K-index are used to characterize the magnitude of geomagnetic storms. These storms can affect our power grids, space operations and radio signals.

Alaska Railway Northern Lights Tour – Appendix A

Background

The Alaska Railroad offers winter vacation packages and one of them is called the Aurora Package. It is a 6-night itinerary aboard the winter train traveling from Anchorage to Fairbanks. The tops of the train cars are all glass so that travelers can have a spectacular view of the Northern Lights. The problem statement for this project was to recommend the times for promotions, since some airlines have been starting to offer percentages off tickets based on the kp index. The Alaska Railroad could also offer promotions during the active sun cycle and solar winds.

Data

The data for this analysis was pulled from University of Alaska Fairbanks Geophysical Institute website using a html get request. Then I used the Beautiful Soup library to scrape and parse the HTML data. The results were placed into a dictionary with the variables 'predicted_time and kp' then converted to a pandas dataframe. Predicted_time was showing as an object and not a datetime, so I switched it to a datatime and then generated new columns by month, day and year. This format would make it easier to see any relationships within the data.

|	predicted_time	kp	year	month	day
0	2020-09-22	2	2020	9	22
1	2020-09-23	3	2020	9	23
2	2020-09-24	4	2020	9	24
3	2020-09-25	4	2020	9	25
4	2020-09-26	4	20		

New DataFrame – Appendix B

Methods

For the analysis, I followed the Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM is a process encompassing six steps that guide us to develop and implement models from data. These steps are, in order, business understanding, data understanding, data preparation, modeling, evaluation, and deployment. During the business understanding portion, the problem was identified, and a solution was purposed.

For the data understanding step, many references were read and viewed to aid in the understanding of the data set, especially the space weather terminology and kp index. The background information was attributed to The Alaska Railroad and the Fairbanks Geophysical Institute. During the data preparation stage, there were no missing items identified and it took a lot of trail and error to scrape the information that was required to make an analysis. Also, in this step of CRISP-DM the data was explored further with plotting the value counts of the variables with multiple plot libraries.

Python was used to gather and explore the data. The most insightful relationships are below. What the data tells me, it is easy to spot the sun cycle activity during the latter part of the month. The highest kp index reached within the past six months was a level of 6. Anything above an index of 4 has the potential to display the aurora. Also, there are a lot calmer nights than there are active ones.
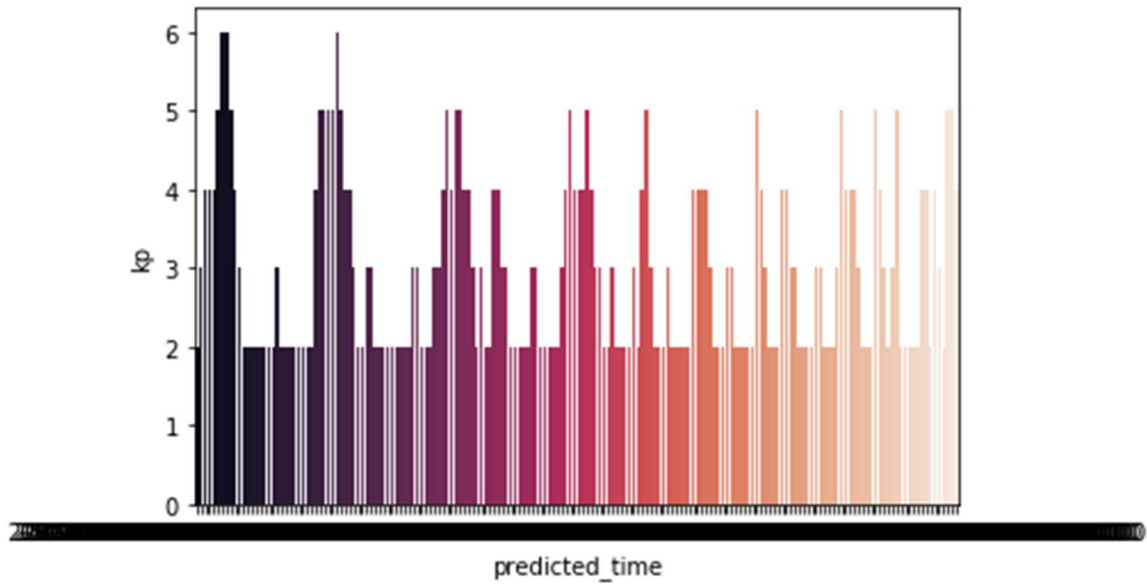
Plot predicted_time and kp Index – Appendix C
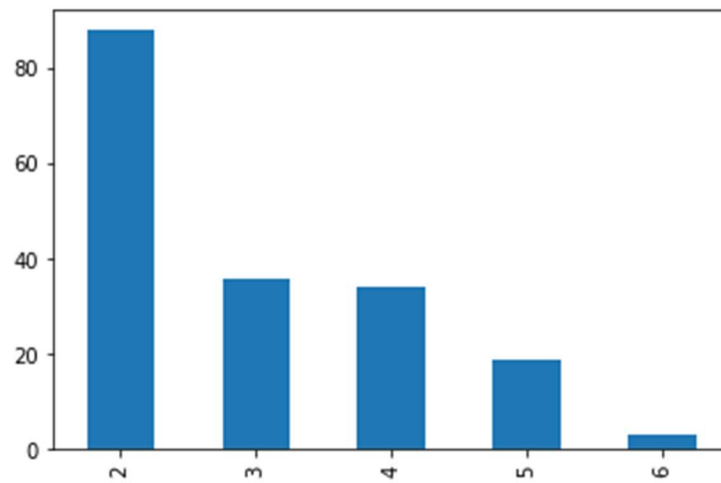
Plot grouped by kp index – Appendix D

Above we see within a 180-day time period there are 80 days of no aurora activity with a little over 20 days of probability to see the aurora.

Results

Once I completed data analysis, I tried to use a recommender model to recommend the days for a promotion. I realized that it was not the best model to use for this data. There were no reviews available for the product, a winter train ride. I research regression models for promotions and decided to change directions. The dataset was trained, tested, and split with test size at 33 percent with x variables being 'month', 'day', 'year' and y being 'kp'. Afterward, it was normalized and then fitted. The trained data was modeled with Linear Regression and predict was used on both trained and tested data. The percentage of accuracy was low, so I tried the Ordinary Least Squares method. The results yielded that the best months for viewing the aurora borealis were 22 to 26. Which looks like it falls on the sun's natural sunspot cycle. The best months were between September to January.
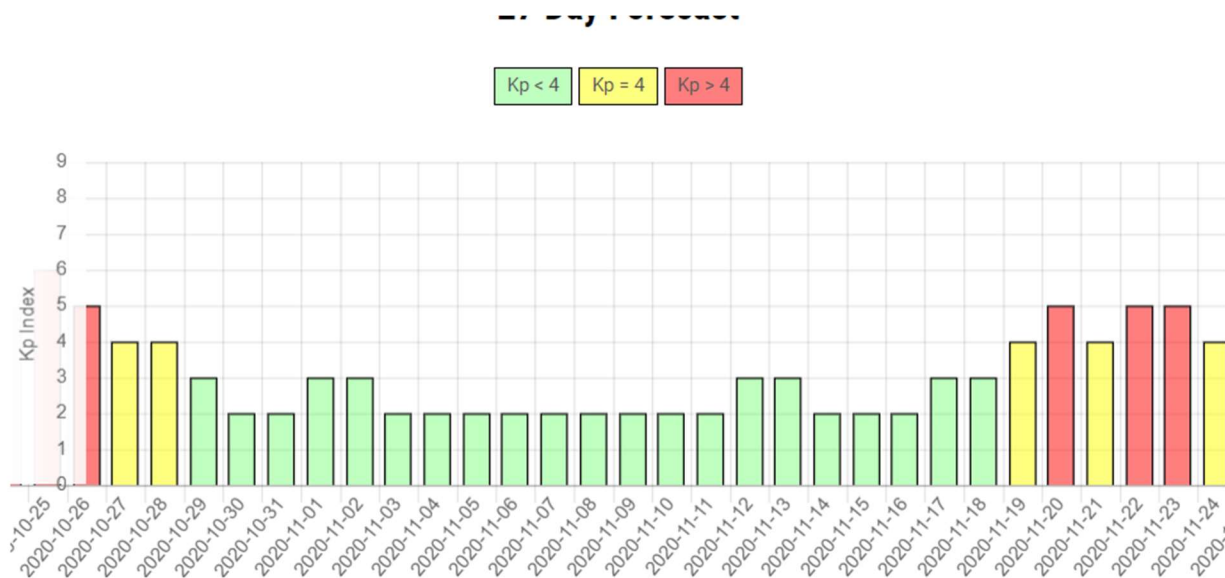
Conclusion

After understanding which model was best to use for this problem statement. I would recommend that the Alaska Railroad focus on the months of September to January and the days of 22 to 26 to promote a winter vacation package.

Acknowledgements

Alaska Railroad, University of Alaska Fairbanks Geophysical Institute, Bellevue University

References:

1. Northern Lights Centre, What are the Northern Lights?, Retrieved from:
   https://www.northernlightscentre.ca/northernlights.html

2. National Oceanic and Atomospheric Administration, Space Weather Prediction Center.
   Swpc.noaa.gov/noaa-scales-explanation. This is a way to communicate current and
   future space weather conditions and the possible effects on people and systems.
   https://www.swpc.noaa.gov/

3. University of Alaska Fairbanks, Geophysical Institute, Aurora Forecast. This is a daily
   forecast of geomagnetic activity on a KP Index.
   https://www.gi.alaska.edu/monitors/aurora-forecast

4. Science on a Sphere, National Oceanic and Atmospheric Administration. Earth at night,
   a data set that was recorded by the Defense Meteorological Satellite Program using polar
   orbiting satellites that provide global cover. https://sos.noaa.gov/datasets/nighttime-
   lights/

5. Data.gov, Solar Features – Solar Flares, data on radiation emitted from sunspots National
   Oceanic and Atmospheric Administration. Solar activity causes the kp index of aurora to
   increase. Dataset is available for download.

6. Space Weather Live, Real time auroral and solar activity. Archived data available of
   geomagnetic activity, solar winds, flare and kp index.

7. Deal Alert, The Points Guy, How to use promotions to see the northern lights for up to
   35 percent off. Fan, K., (2020), Deals Alert, Retrieved from:
   https://thepointsguy.com/deals/deal-   alert-visit-alaska-to-see-the-northern-lights-for-up-
   to-35-off/

Historical references of best times to see the lights.

8. The Northern Light Centre, Northern Lights, Historical data to the Geographic Institute.
     https://www.northernlightscentre.ca/northernlights.html

9. Notifications Archive of Aurora KP indexes. https://auroranotify.com/data/archives/

10. Marketing Dive, Alaska Airlines ties Airfare discounts to Northern Lights., Williams, R.,
     (2020), Retrieved from: https://www.marketingdive.com/news/alaska-airlines-ties-airfare-discounts-to- northern-lights/570365/

11. Making New Year's Travel Resolutions more Affordable Through the Power of Science,
     (2020), The
     higher the kp index the deeper the discount. Retrieved from:
     https://finance.yahoo.com/news/alaska-airlines-harnesses-northern-lights-120000593.html
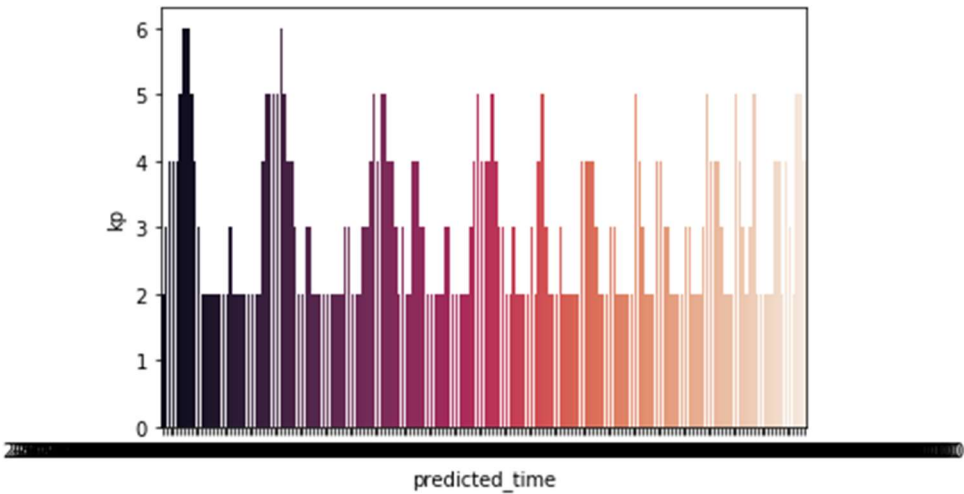
Appendix
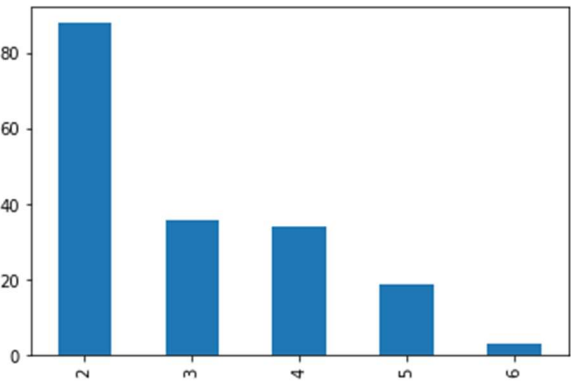
Appendix A:

Alaska Railway Northern Lights Tour

Appendix B: Final dataset

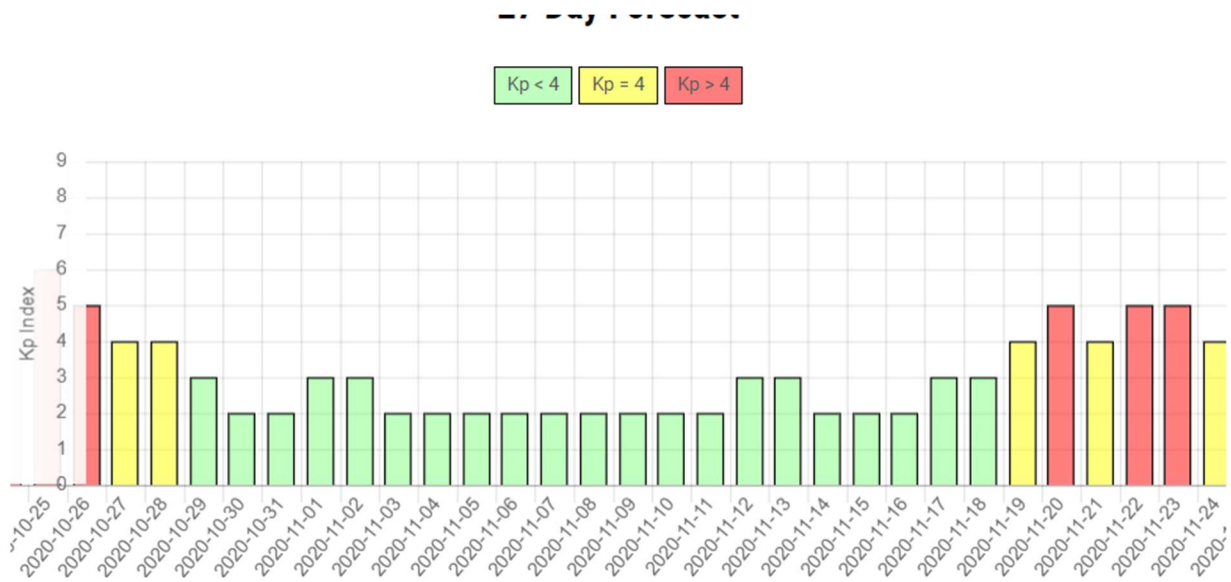|	predicted_time	kp	year	month	day
0	2020-09-22	2	2020	9	22
1	2020-09-23	3	2020	9	23
2	2020-09-24	4	2020	9	24
3	2020-09-25	4	2020	9	25
4	2020-09-26	4	20		

Appendix C: Plot of predicted_time and kp

Appendix D: Grouped by kp

Appendix E: predicted_time and kp

Ten Questions:

When is the best month for viewing the aurora?

When is the best day for viewing the aurora?

Why did you choose this project?

Did anything surprise you in the data analysis?

What did not go as planned?

What steps did you take to overcome those obstacles?

Are the results what you expected?

What did you lessons did you learn in this project?

Would you use something like this for other promotions?

What went right during this project?