# Predictive Modeling for Music Genres Using Machine Learning

## Thompson Morgan

## Abstract

Spotify has a web API that allows users to retrieve data from different content. The information available for tracks includes the audio features, which are calculated by Spotify using audio analysis models. One of the main techniques of machine learning is called classification, which is predicting what class or group an instance of data should be labeled. The goal of this project is to determine which of the selected machine learning algorithms is best at predicting a song's music genre when using the audio features, and to discover which audio features are the most influential for the predictions. The three machine learning algorithms I used are K-Nearest Neighbors, Logistic Regression, and Random Forests. The Random Forest model had the highest accuracy of 56.24%, which was then followed by the K-Nearest Neighbors model with 46.96%. The Logistic Regression model had the lowest accuracy of 45.71%. The Random Forest model calculated that tempo, speechiness, and danceability were the most important features for making a prediction.

## Introduction

Spotify is known as the largest audio streaming platform in the world, and it contains over 100 million different tracks [1]. Spotify has a web API that allows users to retrieve data from different content such as artists, playlists, and tracks. The information available for tracks includes the audio features, which are calculated by Spotify using audio analysis models. The 12 audio features are danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, and duration. Spotify's web API gives definitions for these audio features [2]. Danceability measures how suitable a track is for dancing based on a combination of musical elements. Energy measures the intensity and activity of a track. Key is the key the track is in, mapped to integers using standard pitch class notation. Loudness is the amplitude given in decibels, and it is averaged across the entire track. Mode is whether a track is in major or minor modality, with major being equal to one and minor being equal to zero. Speechiness is a measure of the presence of spoken words. Instrumentalness predicts whether a track contains vocals or not. Liveness detects the presence of a live audience in a track. Valence is a measure of how positive the music of a track is. Tempo is the estimated speed of a track measured in beats per minute. Lastly, the duration is how long a track is, and it is measured in milliseconds.

Machine learning is at the intersection of computer science, engineering, and statistics, and can be summed up as turning data into information [3]. One of the main techniques of machine learning is called classification. Classification is predicting what class or group an instance of data should be labeled [4]. There are many different algorithms available for classification, but the three this project focuses on are K-Nearest Neighbors, Logistic Regression, and Random Forests. K-Nearest Neighbors takes the "k" closest points to an input and predicts the input's class based on the majority class of those points [5]. Logistic regression takes the input data and develops an equation to use for classification [6]. Random forest is a method that combines the output of multiple decision trees to make a prediction [7].

The goal of this project is to determine which of the selected machine learning algorithms is best at predicting a song's music genre when using the audio features, and to discover which audio features are the most influential for the predictions.

## Methods

I completed this project using Python. I used the Pandas and NumPy packages for handling the data, and I used the Matplotlib and Seaborn packages for data visualization. I also used Scikit-Learn packages for the machine learning algorithms and the performance evaluations. The code for this project is available on my GitHub page [8]. The dataset can also be found in the "tidytuesday" repository on the "rfordatascience" GitHub page [9].

I began this project by cleaning and exploring the dataset. To clean the data, I checked for any missing or duplicate values. I also checked for any spelling errors in the genre column. To explore the data, I generated a count plot to visualize the total number of tracks for each of the genres in the dataset. I also created density plots for each genre of every audio feature. I did this to study the relationships between the genres and each of the audio features.

For each machine learning algorithm, I split the data into training and testing datasets with a ratio of 80:20, and I used the same random state seed each time to ensure consistency. I normalized the audio features data before using it to train and test the models. I generated a confusion matrix and classification report for each algorithm to assess the results.

The first machine learning algorithm I used was K-Nearest Neighbors. I used the elbow method to find the optimal value of k. To do this, I calculated the error rates of k-NN models that used values of k from two to 20. Next, I visualized the results and found the value for k where the line graph started to level off.

The second machine learning algorithm I used was Logistic Regression. I added a column of ones to the data frame of audio features to represent the bias, which is a standard practice when using logistic regression. I completed this task with the Statsmodels package.

The third machine learning algorithm I used was Random Forests. I created a data frame containing the importance of each audio feature when making a prediction, which is calculated when generating the Random Forest model.

## Results

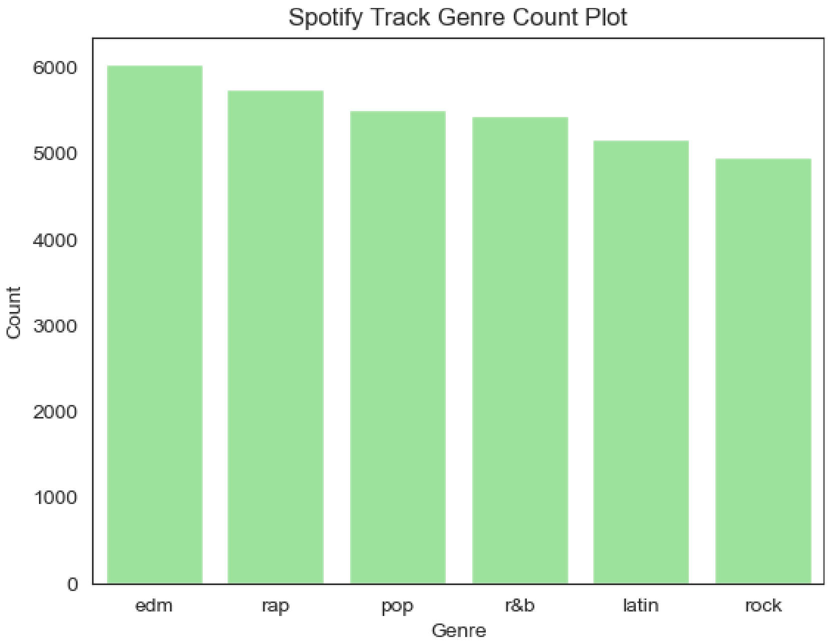


***Figure 1.*** *The total number of tracks for each of the six genres in the dataset.*

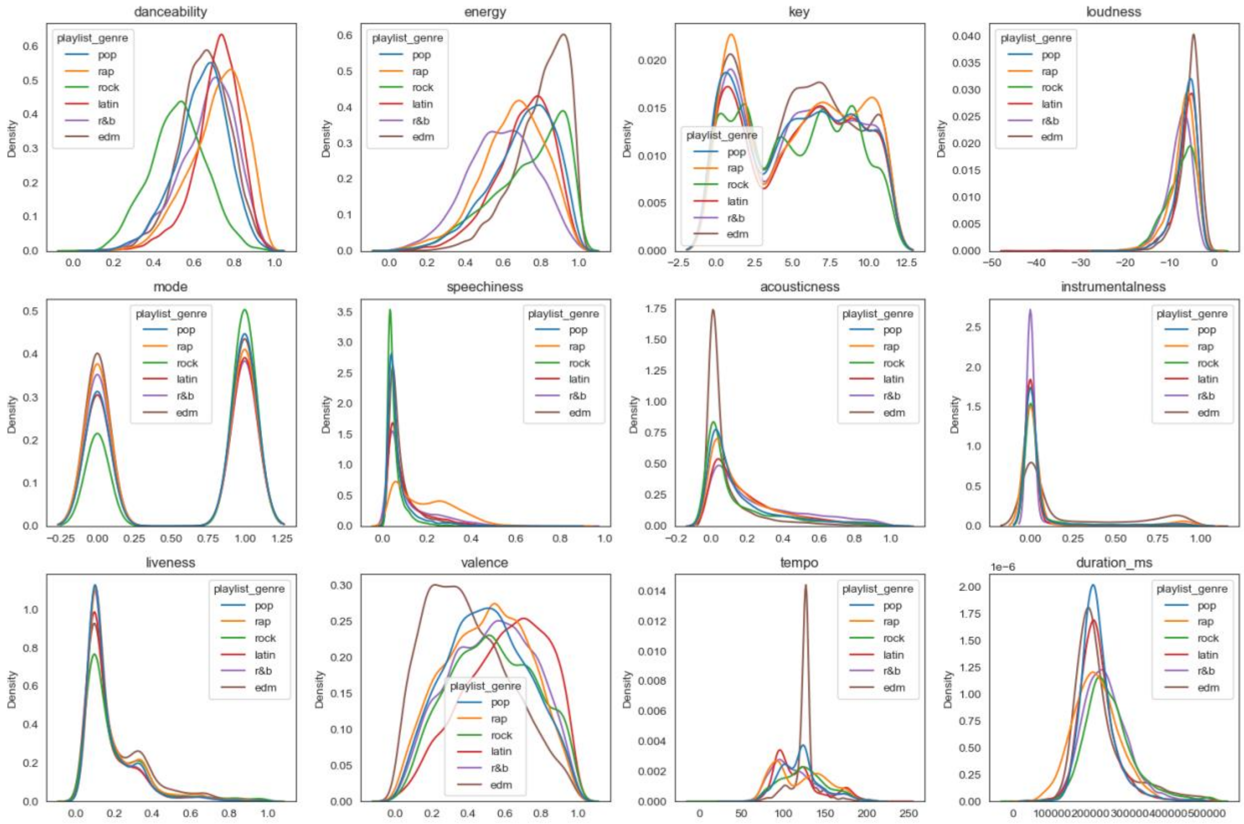


*__Figure 2.__ Density plots of every audio feature for each of the music genres.*

The K-Nearest Neighbors model was ranked second out of the three machine learning algorithms. The model accurately predicted the genre of the track 46.96% of the time. The line graph shown in figure three demonstrates how I determined the optimal value of k to be seven using the elbow method. The classification report shown in table one provides the main performance metrics of the model. The confusion matrix shown in figure four compares the true and predicted labels for each genre.

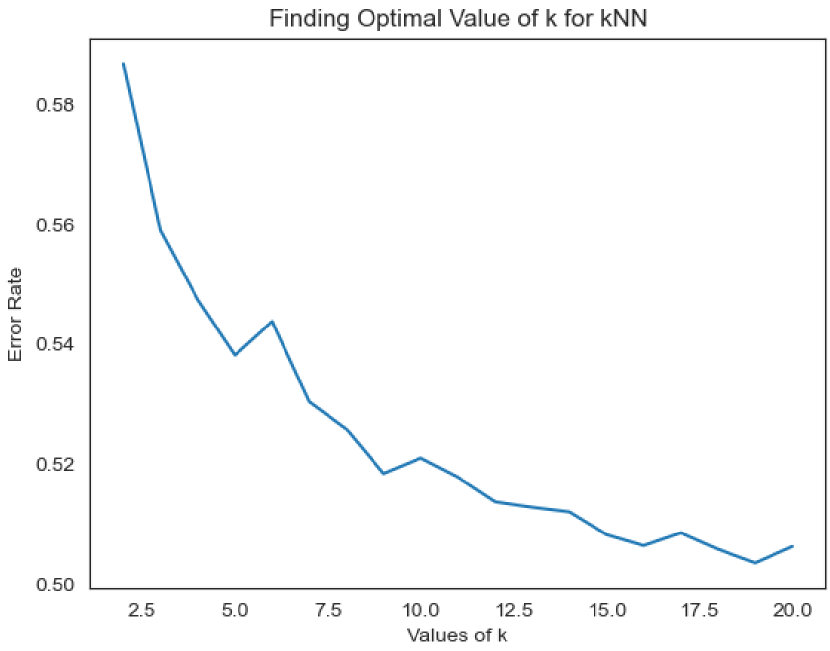
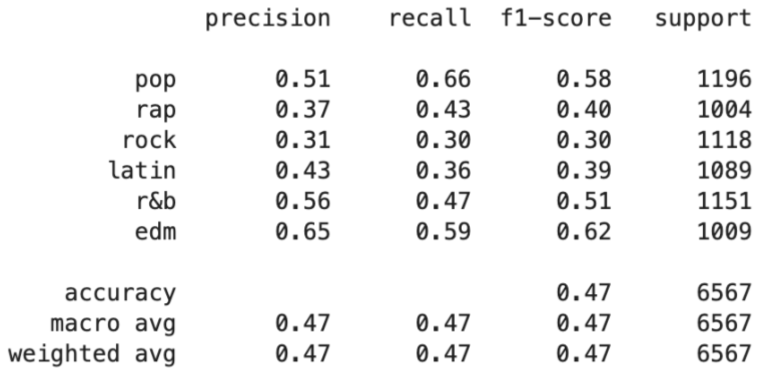


***Figure 3.*** *A line graph demonstrating the error rate for values of k from two to 20.*



```
              precision    recall  f1-score   support

         pop       0.51      0.66      0.58      1196
         rap       0.37      0.43      0.40      1004
        rock       0.31      0.30      0.30      1118
       latin       0.43      0.36      0.39      1089
         r&b       0.56      0.47      0.51      1151
         edm       0.65      0.59      0.62      1009

    accuracy                           0.47      6567
   macro avg       0.47      0.47      0.47      6567
weighted avg       0.47      0.47      0.47      6567
```

***Table 1.*** *The classification report generated for the K-Nearest Neighbors model.*

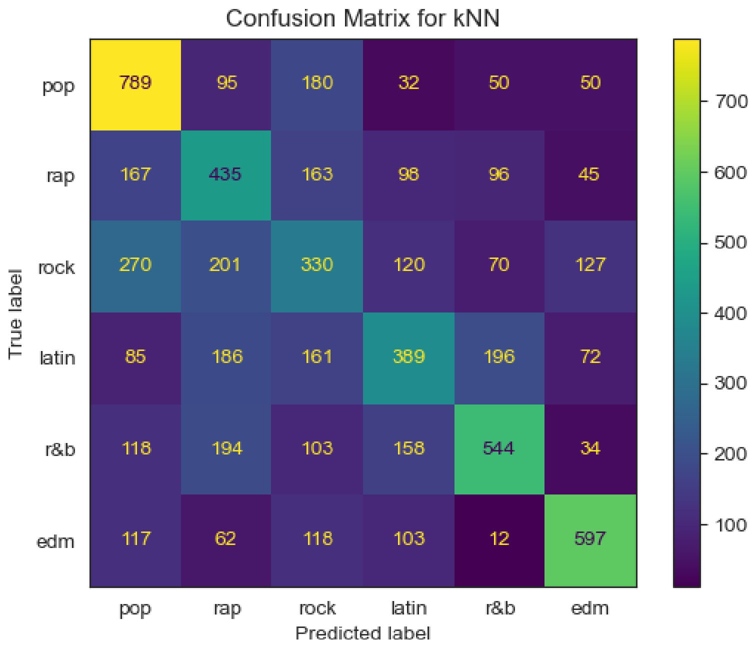
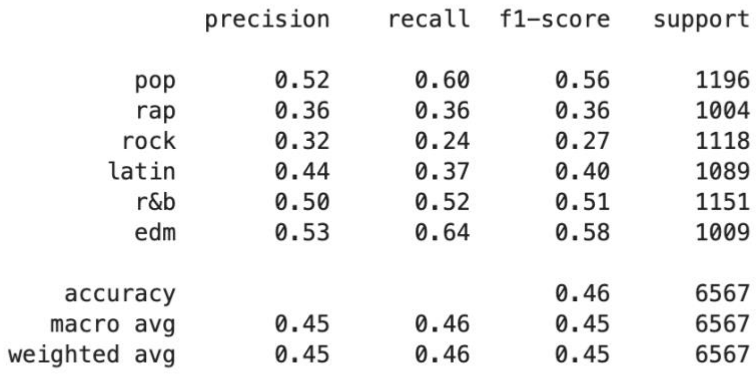


***Figure 4.*** *A confusion matrix showcasing the results of the K-Nearest Neighbors model.*

The Logistic Regression model was ranked third out of the three algorithms. The model accurately predicted the genre of the track 45.71% of the time. The classification report shown in table two provides the main performance metrics of the model. The confusion matrix shown in figure five compares the true and predicted labels for each genre.



```
              precision    recall  f1-score   support

         pop       0.52      0.60      0.56      1196
         rap       0.36      0.36      0.36      1004
        rock       0.32      0.24      0.27      1118
       latin       0.44      0.37      0.40      1089
         r&b       0.50      0.52      0.51      1151
         edm       0.53      0.64      0.58      1009

    accuracy                           0.46      6567
   macro avg       0.45      0.46      0.45      6567
weighted avg       0.45      0.46      0.45      6567
```

***Table 2.*** *The classification report generated for the Logistic Regression model.*

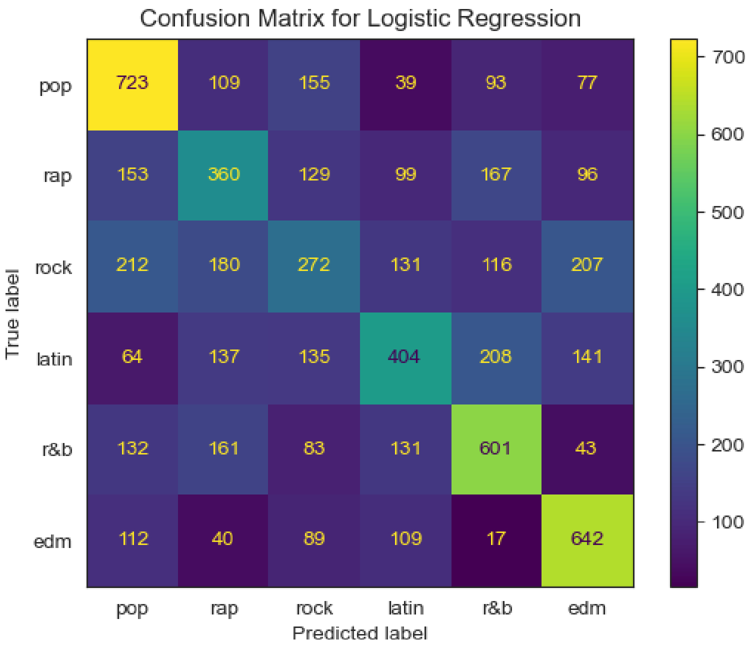
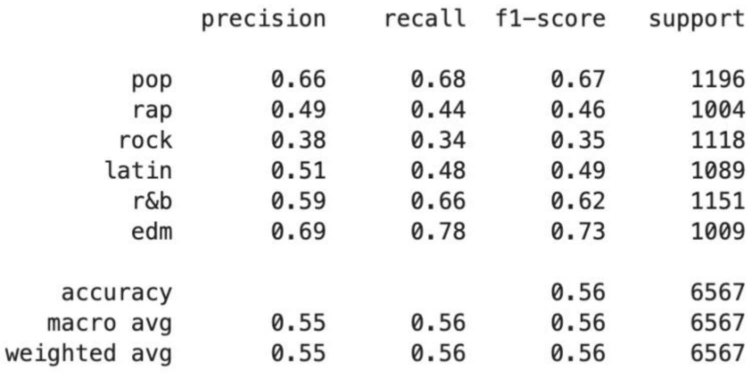


***Figure 5.*** *A confusion matrix showcasing the results of the Logistic Regression model.*

The Random Forest model was ranked first out of the three algorithms. The model accurately predicted the genre of the track 56.24% of the time. The classification report shown in table three provides the main performance metrics of the model. The confusion matrix shown in figure six compares the true and predicted labels for each genre. Table four shows the importance of each audio feature for making predictions.


```
              precision    recall  f1-score   support

         pop       0.66      0.68      0.67      1196
         rap       0.49      0.44      0.46      1004
        rock       0.38      0.34      0.35      1118
       latin       0.51      0.48      0.49      1089
         r&b       0.59      0.66      0.62      1151
         edm       0.69      0.78      0.73      1009

    accuracy                           0.56      6567
   macro avg       0.55      0.56      0.56      6567
weighted avg       0.55      0.56      0.56      6567
```

***Table 3.*** *The classification report generated for the Random Forest model.*

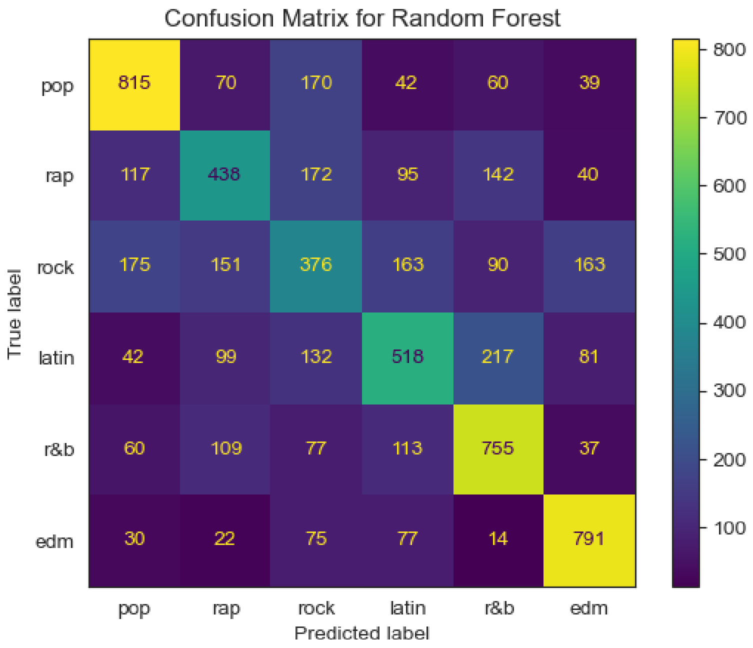


***Figure 6.*** *A confusion matrix showcasing the results of the Random Forest model.*

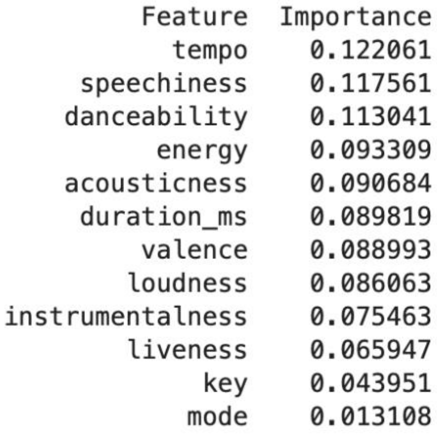

```
           Feature  Importance
             tempo    0.122061
       speechiness    0.117561
      danceability    0.113041
            energy    0.093309
      acousticness    0.090684
       duration_ms    0.089819
           valence    0.088993
          loudness    0.086063
  instrumentalness    0.075463
          liveness    0.065947
               key    0.043951
              mode    0.013108
```

***Table 4.*** *The importance of each audio feature for making predictions calculated by the Random Forest model.*

## Discussion

The density plots illustrated in figure two help give insight into the relationships between the genres and each of the audio features. Danceability, speechiness, and tempo appear to have the most distinct density plots for each genre. Key, mode, and liveness appear to have similar density plots for each genre. This is supported by the audio feature importance presented in table four, as tempo is ranked as the most important and mode is ranked as the least for making a prediction.

The Random Forest model had the highest accuracy of 56.24%, which was then followed by the K-Nearest Neighbors model with 46.96%. The Logistic Regression model had the lowest accuracy of 45.71%.

The K-Nearest Neighbors and Logistic Regression models had very similar accuracies, but each was better at predicting different genres than the other. The k-NN model accurately predicted pop, rap, and rock tracks more often. Conversely, the Logistic Regression model accurately predicted latin, R&B, and EDM tracks more often. The Random Forest model was better than the other models at all six genres as well.

In the future, the same data can be used with new machine learning algorithms to compare the new results to those I obtained during this project. A few other methods that could be implemented are K-Means Clustering, Naïve Bayes, or neural networks.

## Acknowledgements

I would like to start off by thanking the department of Data Science and Strategic Analytics at Stockton University. During my time in the program, I learned so much and found new confidence in my programming and data science skills, and this was thanks to all the wonderful professors that make the curriculum both informative and engaging. I would also like to thank my parents, my brothers Lennon and Holden, my girlfriend Gabriella, and my cat Lilo. I would not have been able to go back to university without their love and support. I would like to end by also thanking the Data Science Learning Community for hosting Tidy Tuesdays and making their datasets available for public use.

## References

[1] Singh, Shubham. (2025, May 16). *Spotify User Statistics (2025) – Subscribers and Demographics Data*. Demandsage. https://www.demandsage.com/spotify-stats/

[2] https://developer.spotify.com/documentation/web-api/reference/get-audio-features

[3] Harrington, P. (2012). *Machine Learning in Action*, p. 5. Manning Publications Co.

[4] Harrington, P. (2012). *Machine Learning in Action*, p. 10. Manning Publications Co.

[5] (2025, July 23). *K-Nearest Neighbor(KNN) Algorithm*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/k-nearest-neighbours/

[6] Harrington, P. (2012). *Machine Learning in Action*, p. 83-84. Manning Publications Co.

[7] *What is random forest?*. IBM. https://www.ibm.com/think/topics/random-forest

[8] https://github.com/tjmorgan462/spotify-ml

[9] https://github.com/rfordatascience/tidytuesday/blob/main/data/2020/2020-01-21/readme.md