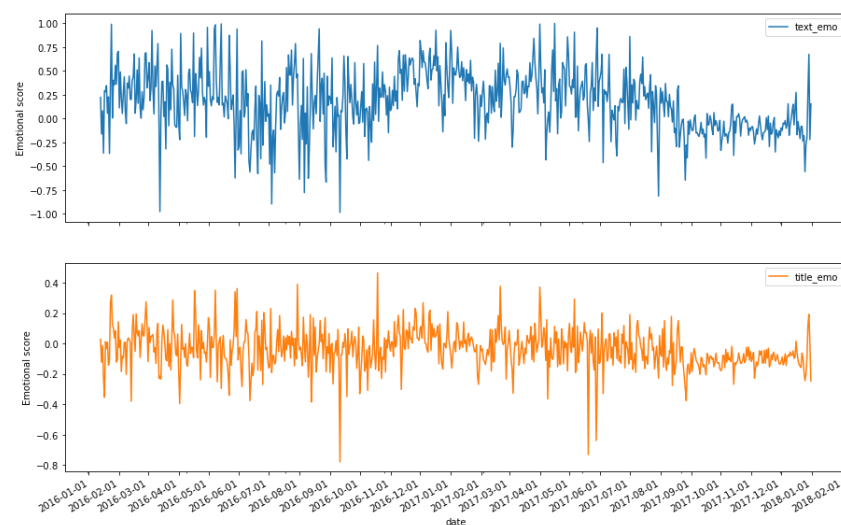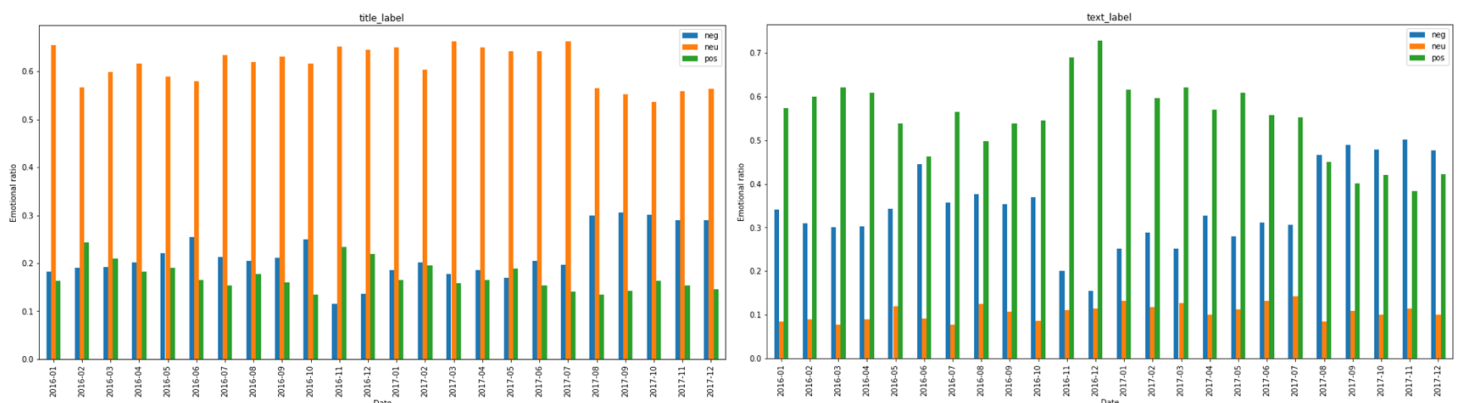# Project

The project data comes from "fake and real news dataset" in Kaggle. In the exploratory data analysis, it mainly analyzes the emotion of true and false news. Firstly, read the true and false data respectively, then analyze the emotion of each news through the relevant modules of NLTK, and finally draw the line chart, column chart and pie chart. In order to draw the pie chart, the scores of the affective analysis are grouped, and those greater than 0.33 are set as positive sentences and those less than -0.33 are set as negative sentences.

1. Exploratory data analysis of true news

The first is the line chart of the real news, the blue is the line chart of the news content, the orange is the line chart of the news title, and the horizontal axis is the month.
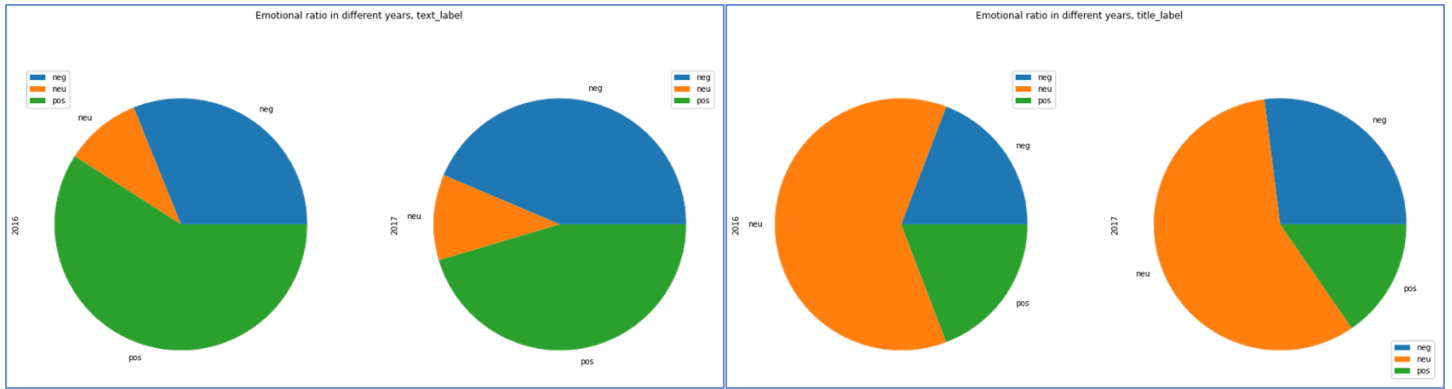

As can be seen from the figure, this data set is the real news from 2016 to 2018. At the same time, it can be seen that the emotional fluctuation of news from September to December 2017 is relatively small, indicating that there is no big news in this period.

The second chart is the histogram, which is used to analyze the emotion of the title and text respectively. By drawing the histogram, we can see that the proportion of positive, neutral and negative emotion of both title and content is relatively stable. However, the title is neutral and the content is positive.

data sources: https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset
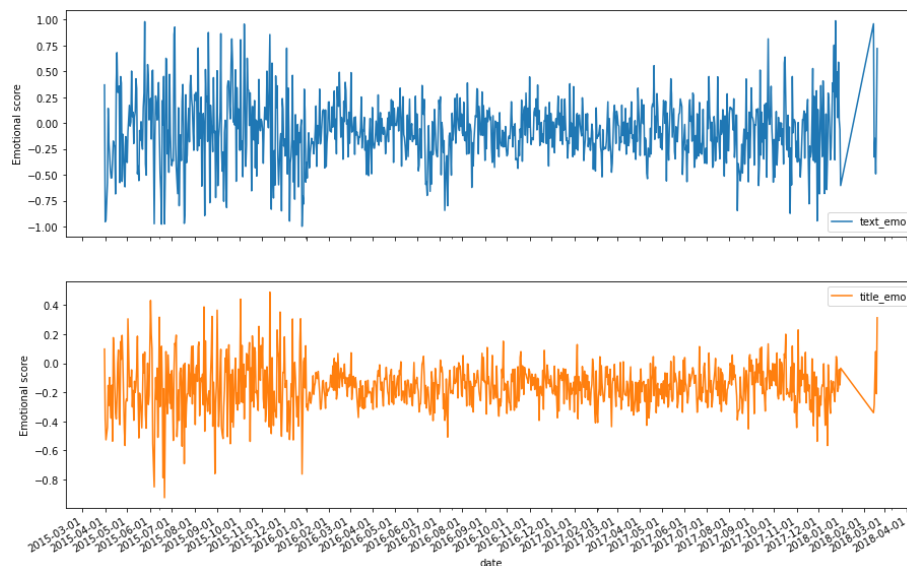
Project

The third chart is the pie chart, which observes the news emotion through the overall proportion. For news content, positive content is the main, but the positive content in 2017 is less than that in 2016. For news headlines, neutral content accounts for the majority, which changes little with the year.

2. Exploratory data analysis of false news

This is the broken line chart of the emotion analysis of fake news. Note that the fluctuation of the title in the fake news is around - 0.2, while the fluctuation of the previous real news is around 0. Maybe this can be used as a judgment basis for the true and false news. At the same time, it is observed that the false news in this data set is from 2015 to 2018, while the true news is from 2016 to 2017.

The last one is the pie chart of fake news. It can be seen that the title of fake news is neutral and negative, while the content of fake news is negative and positive, indicating that fake news tends to have richer feelings in content structure. In terms of time, the emotional proportion of fake news is relatively stable.

data sources: https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset

Project

Emotional ratio in different years, title_label

Emotional ratio in different years, text_label