Time Series Analysis

Nane Mambreyan

Cryptocurrency Dataset

Comparing time series

Comparative Cryptocurrency Metrics Over Time
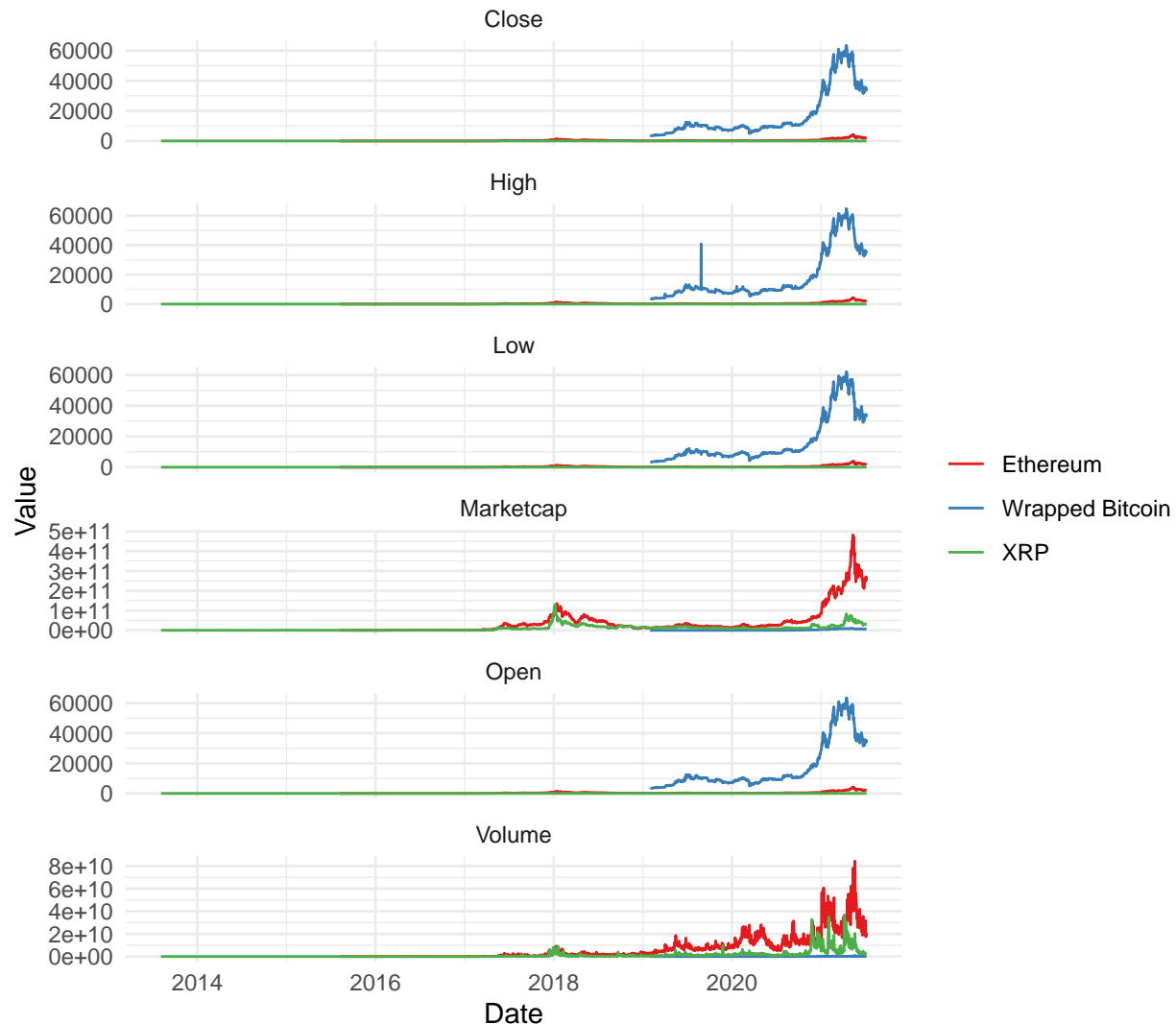
Since the price scales are significantly different for Wrapped Bitcoin and other two (XRP, Ethereum), so it is difficult to make detailed comparisons based on values since the variation for XRP is not clear due to scaling of others. So further inspection is needed. For all plots we notice spikes for 2021, which are more visible for Wrapped Bitcoin. We can also notice relatively lower spikes for period between 2019 and 2020. This period differd only for Marketcap prices which had 1st peak in 2018.

High Prices Over Time: The peaks in the high prices for Ethereum and Wrapped Bitcoin reflect moments of intense buying pressure and market optimism especially in 2021. We can also notice a peak for Wrapped Bitcoin high prices between 2019 and 2020. Sharp spikes might represent periods of speculative trading or reactions to positive news or developments within the cryptocurrency space.

Low Prices Over Time: The lows for Ethereum and Wrapped Bitcoin gradually increase, indicating a rising base price over time. Occasional dips could be the result of market corrections, negative news, or broader economic downturns affecting risk assets like cryptocurrencies for both.

Close Prices Over Time: Ethereum and Wrapped Bitcoin exhibit a strong upward trend in closing prices. This suggests a sustained positive investor sentiment towards these assets. The Wrapped Bitcoin's closing prices appear to be closely aligned with Bitcoin's market movements, given its nature as a representation of Bitcoin on the Ethereum blockchain. XRP's closing price remains relatively stable and lower in comparison, which could indicate a more cautious approach from investors or a different market dynamic affecting its price.

Open Prices Over Time: The opening prices follow a trend similar to the closing prices, with a general upward trajectory for Ethereum and Wrapped Bitcoin. Consistency between open and close prices might indicate a stable market without sudden changes due to overnight news or events.

Volume Over Time: Volume spikes are often associated with price movements, indicating periods of high trading activity which can be driven by news releases, product launches, or various macroeconomic factors. Sustained high volume in conjunction with price increases could validate the trend as being backed by strong investor interest or other factors. In this plot the volume increases gradually for all 3 types, having highest increase for Ethereum, then XRP, then Wrapped Bitcoin.

Marketcap Prices Over Time: This chart mirrors the individual asset price movements over time. We notice sharp increases in market cap, which could be attributed to both price appreciation and the minting of new tokens or coins.

By every measure, Ethereum and Wrapped Bitcoin show robust performance. This could be attributed to the expanding decentralised application ecosystem, growing institutional interest, and the overall positive market conditions surrounding cryptocurrencies over the previous years. When compared to Ethereum and Wrapped Bitcoin, XRP has performed less well in terms of price and market cap growth. This could be a sign of particular difficulties or a lag in adoption. The data indicates that although the cryptocurrency market is expanding, there are times when it is very volatile. Many things, such as developments in technology, regulatory news, market , and changes in investor sentiment, can have an impact on volatility.

Seasonal analysis

Since we are dealing with seasonal analysis, it is reasonable to look at the cost for the end of the season. That characteristic is represented in Close column of our dataset. So we begin with extracting the Dates and closing prices of the cryptocurrency. To be able to create seasonal plots we need time series object, where our seasons are the years. To create a visually interpretable data, we take the frequency to be 12, refering to monthly data.

Wrapped Bitcoin
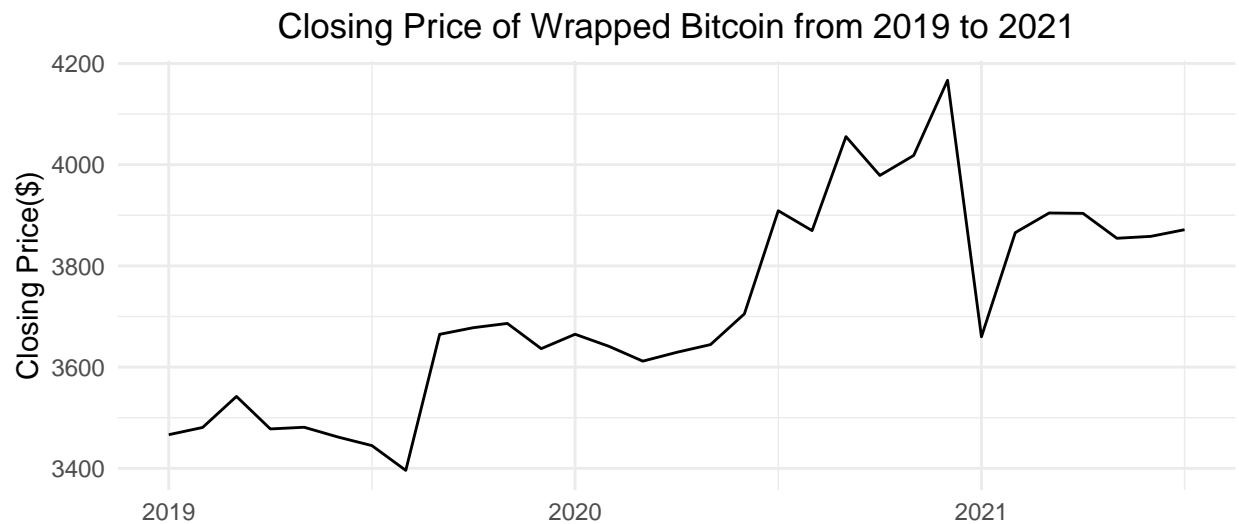
This plot shows the overall trend of the price for this cryptocurrency over years(seasons). We see that in the beginning of the second half of 2019, the price is at its minimal, and it is at its maximal towards the end of 2020. We have drastic decrease of the price in the beginning of 2021. We cannot say anything for sure about cycles and seasonality yet, but the overall trend from 2019 to 2021 is increasing with some periods of drastic decreases.
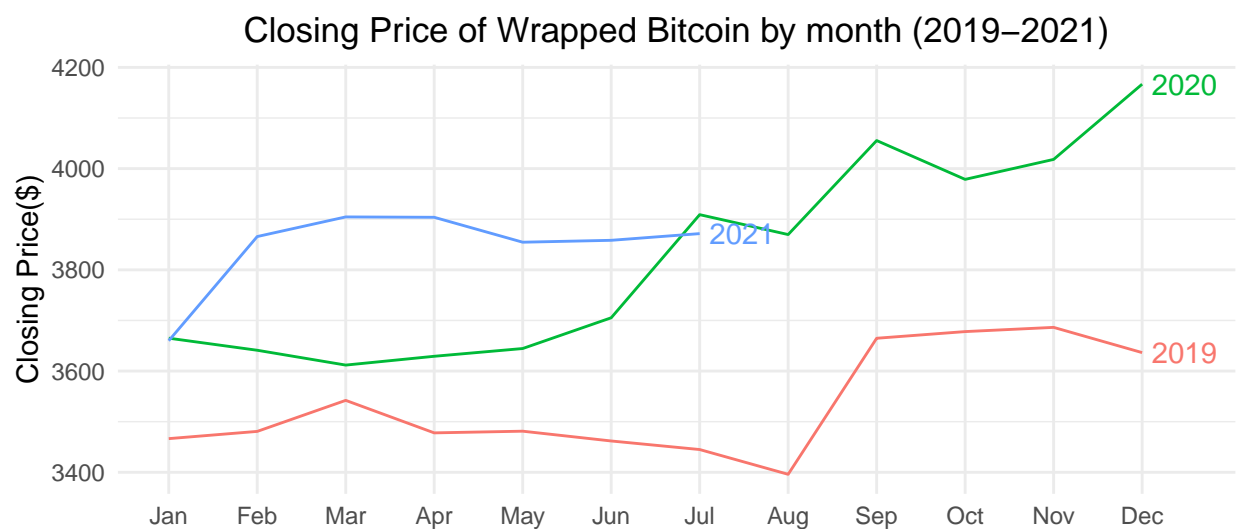
With this plot we already see some seasonality. For 2019 and 2020, we have decrease from Jul to Aug and increase from Aug to Sep, so we can expect something similar for 2021. From 2019 to the end of 2020 the general trend is increasing, especially in 2020, but then right at the beginning of 2021 the price decreases abruptly(the blue line starts with lower value of closing price than the price at which 2020 ends).
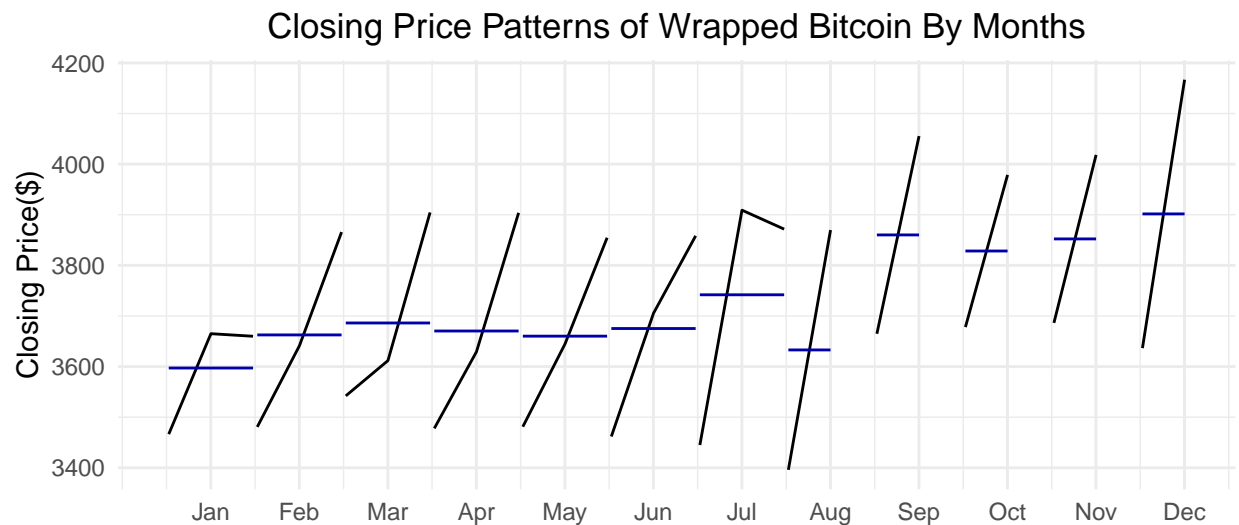
This shows that for the considered years, from Jan to March we have generally increasing trend, then slight decrease till May. After that we see some increasing tendency for two months. As already mentioned above, at the beginning of Aug we see abrupt decrease. After that the price stays more or less stable on max values. The slices are more or less similar, hence we can expect to have some seasonality.
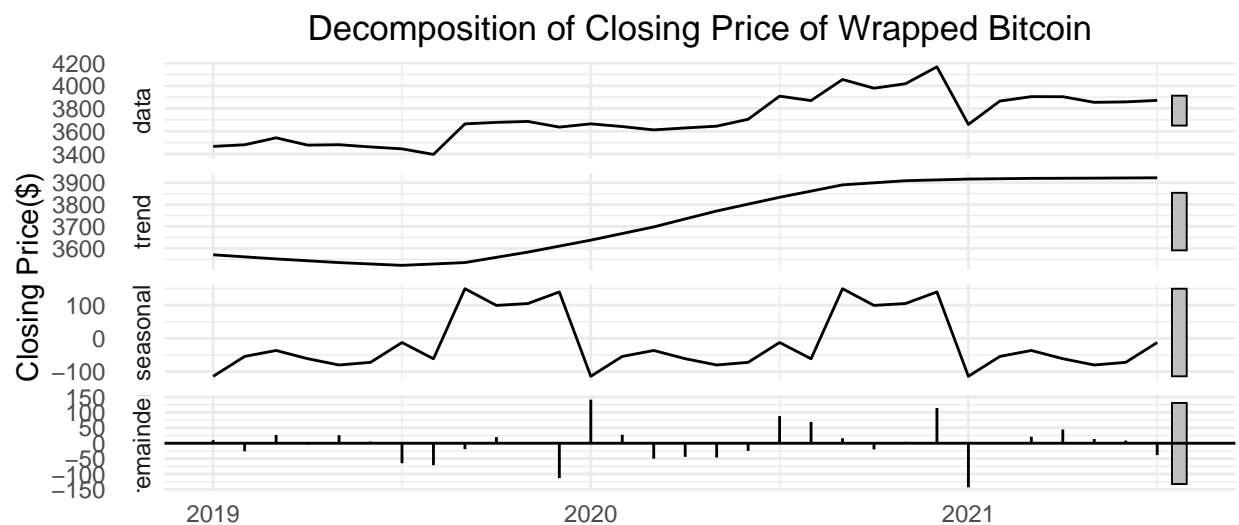
- Trend: We see that overall the price increased within mentioned period

- Seasonality: We see fixed time periods with similar 'visual shapes' on fixed periods within time. This is clear pattern of seasonality.

- Cycles: No cycles on the trend component representing combined trend and cycle patters.

- Remainder: It is random, hence our seasonal and trend analysis can be considered reliable.

4

XRP

Extracting the same information as described above for WrappedBitcoin.

Closing Price of XRP from 2013 to 2021
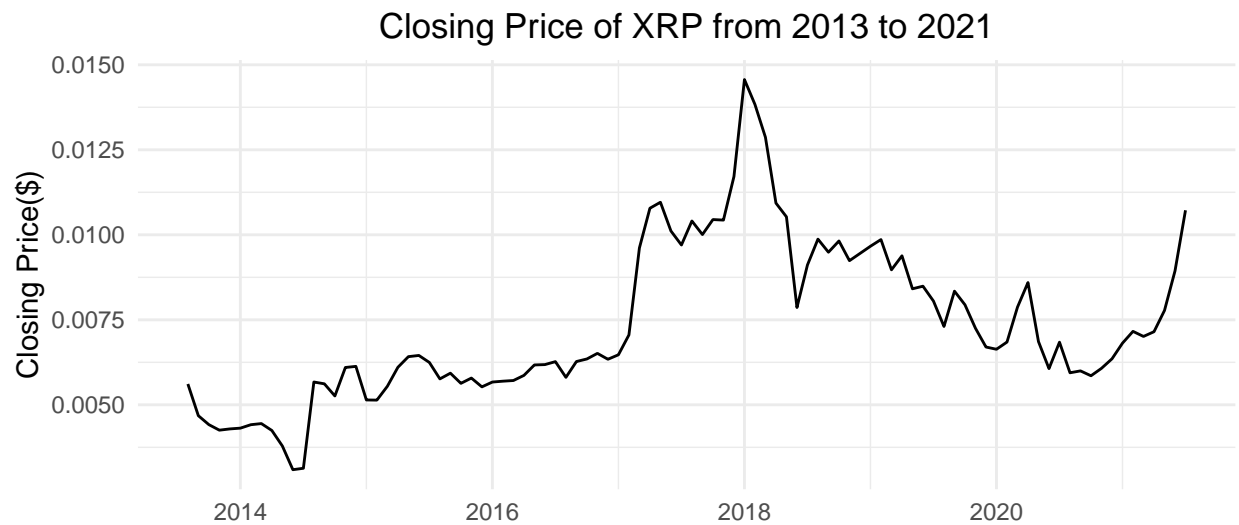
Again there are no specific patterns that can be seen on this plot which shows overall price fluctuations from 2014 to 2020. We see decrease in 2014 and drastical increase in 2017 as well as 2018.

Closing Price of XRP by month (2013–2021)
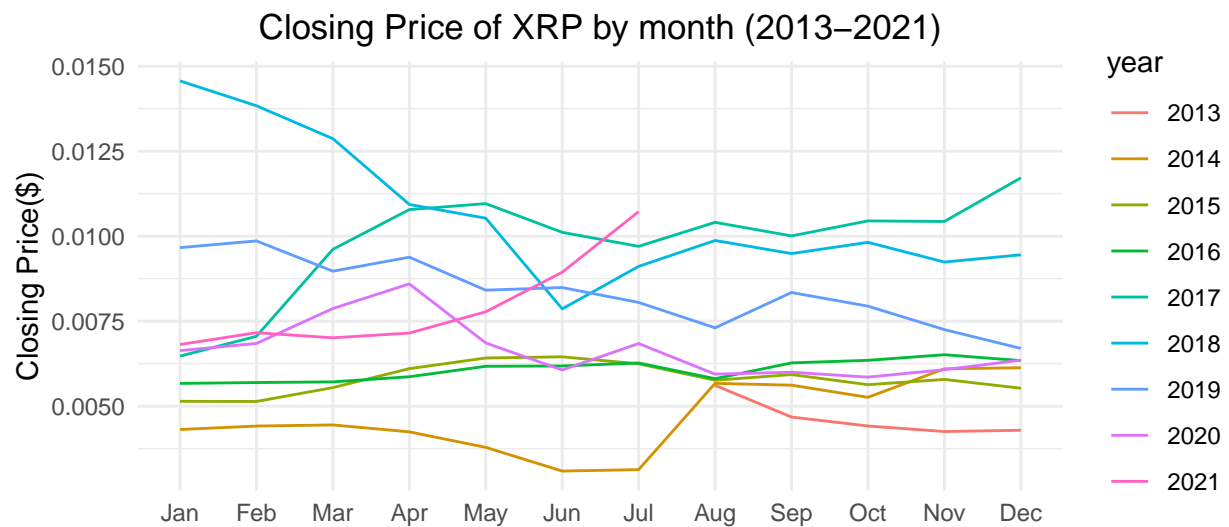
We have 9 years and the picture doesn't have vivid pattern for the depicted lines, hence there is no apparent seasonality for these years and we need further analysis.

Closing Price Patterns of XRP By Months
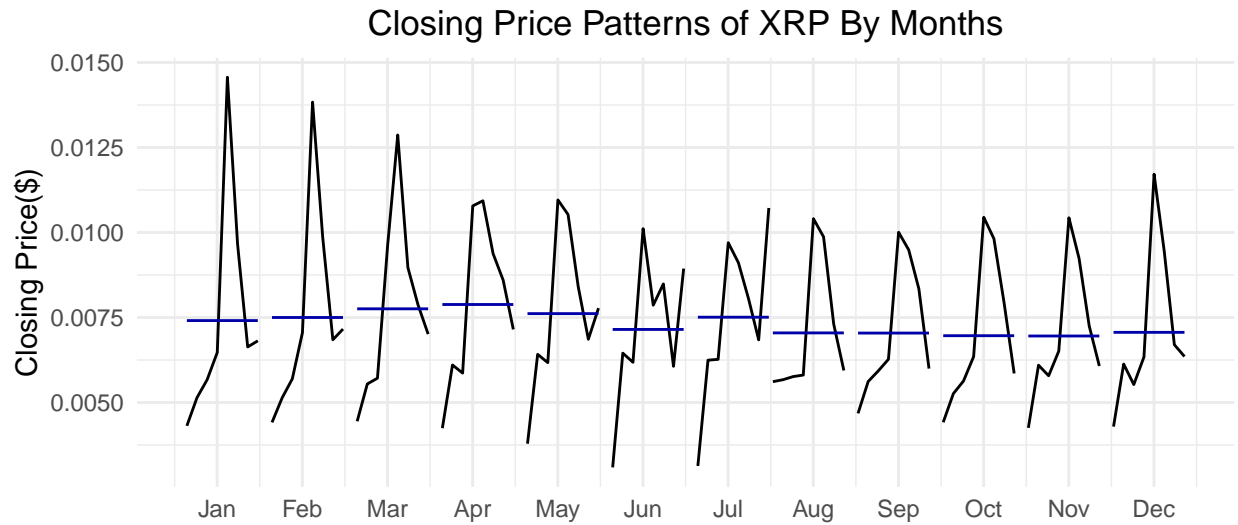
Each vertical "slice" of the plot corresponds to a specific month, and the lines within each slice represent the seasonal component for different years. The consistent presence of peaks within each monthly slice indicates a repeatable seasonal pattern across years.

The blue horizontal line across the plot represent the mean of the seasonal component. A relatively flat line suggests that the average seasonal effect has not changed much over the years, indicating stability in the seasonal pattern.

Decomposition of Closing Price for XRP

- Trend: There is no vivid increasing or decreasing trend but we can see that for example in 2018 the price was on its peak.

- Seasonality: We see fixed time periods with similar 'visual shapes' on fixed periods within time. This is clear pattern of seasonality.

- Cycles: No cycles on the trend component representing combined trend and cycle patters.

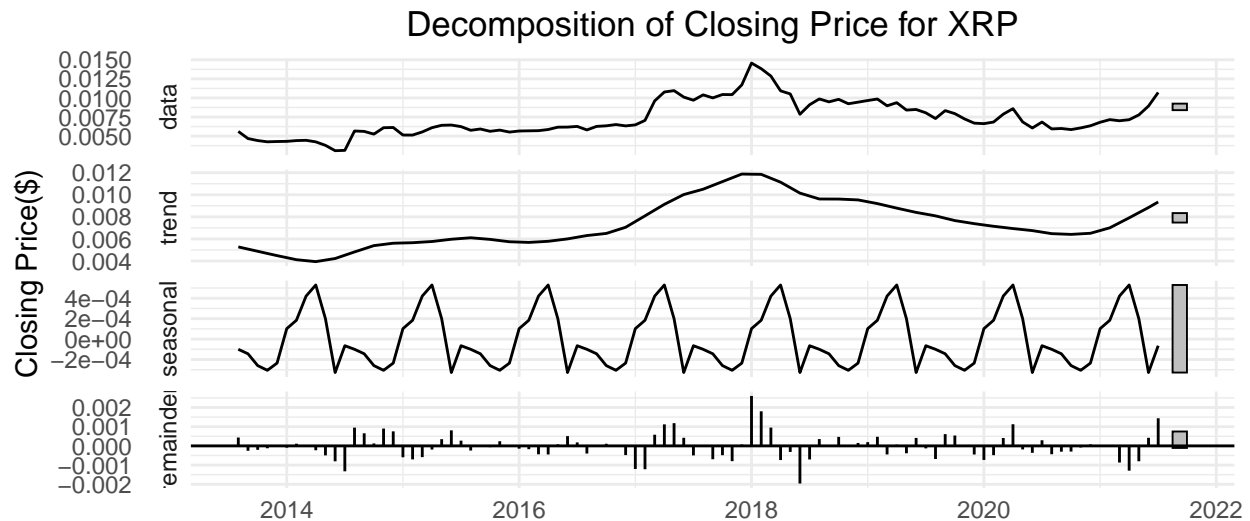- Remainder: It is random, hence our seasonal and trend analysis can be considered reliable.

Ethereum

Closing Price of Ethereum from 2015 to 2021
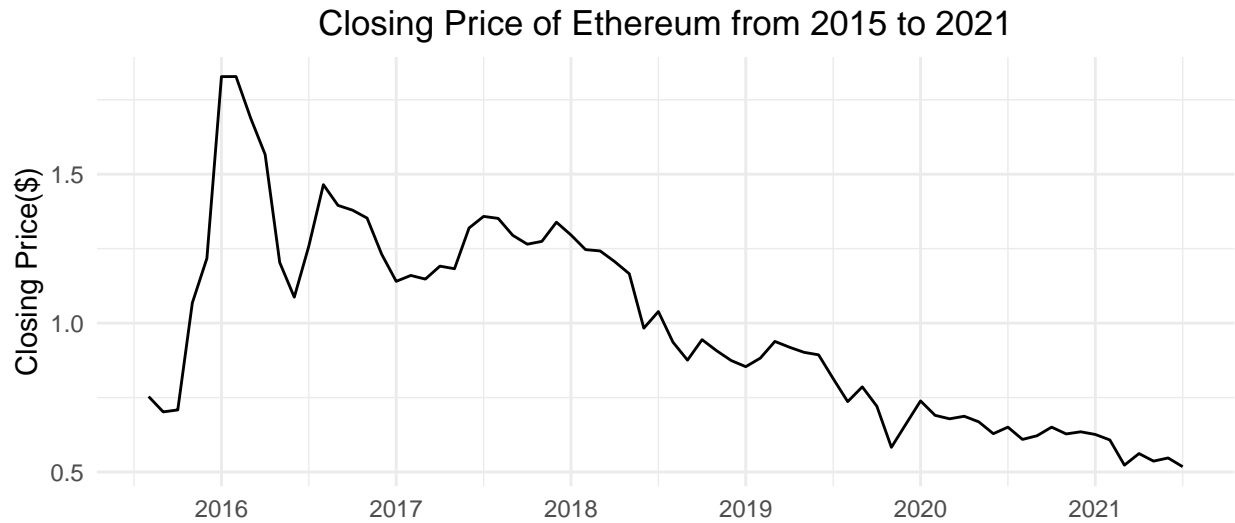
The plot shows emphasized decreasing trend. We generally observe stable decrease with some periods of abrupt decreases. For more explicit patterns we need further analysis:

Closing Price of Ethereum by month (2015–2021)
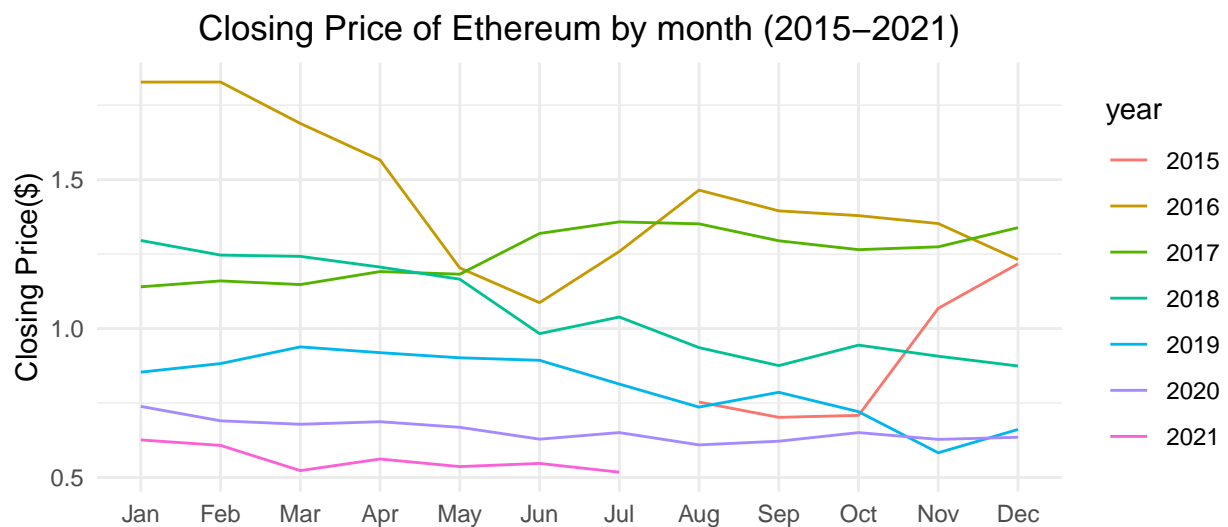
This plot also contributes to the claim of decreasing trend, since we have lowest prices for the last year represented with pink line. Because we have data for multiple years, further patterns are not visible on this plot, so we can proceed to other visualizations.

Closing Price Patterns of XRP By Months
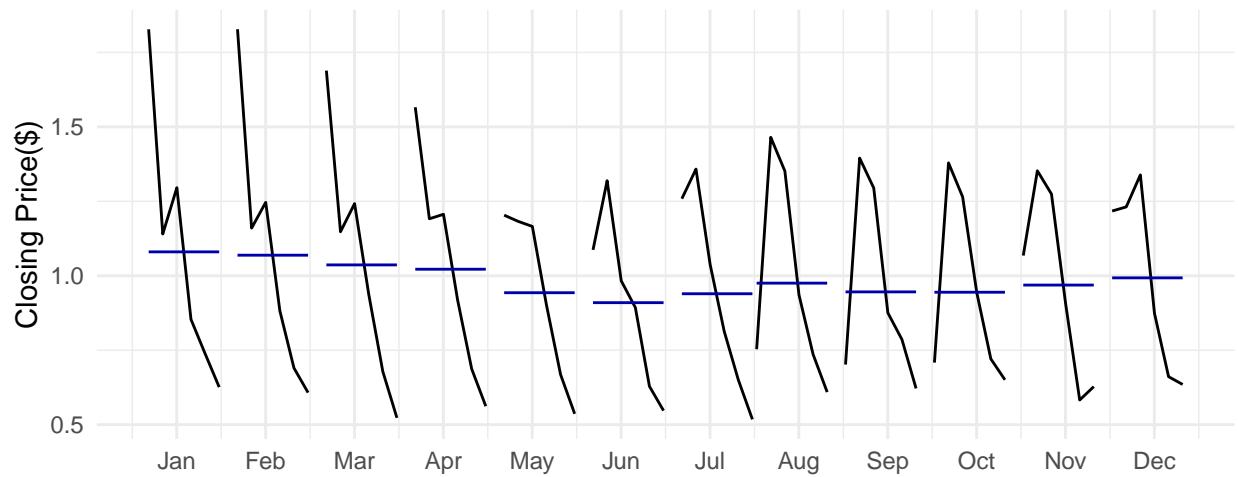
For this plot we already see some seasonality.

The lines within each slice represent the seasonal component for different years. The consistent presence of peaks within each monthly slice indicates a repeatable seasonal pattern across years. This is especially true for periods from Jan to April and Jun to Dec.

The blue horizontal line across the plot representing the mean of the seasonal component also show the seasonal pattern since we see similar increasing and decreasing pattern in them.

Decomposition of Closing Price of XRP
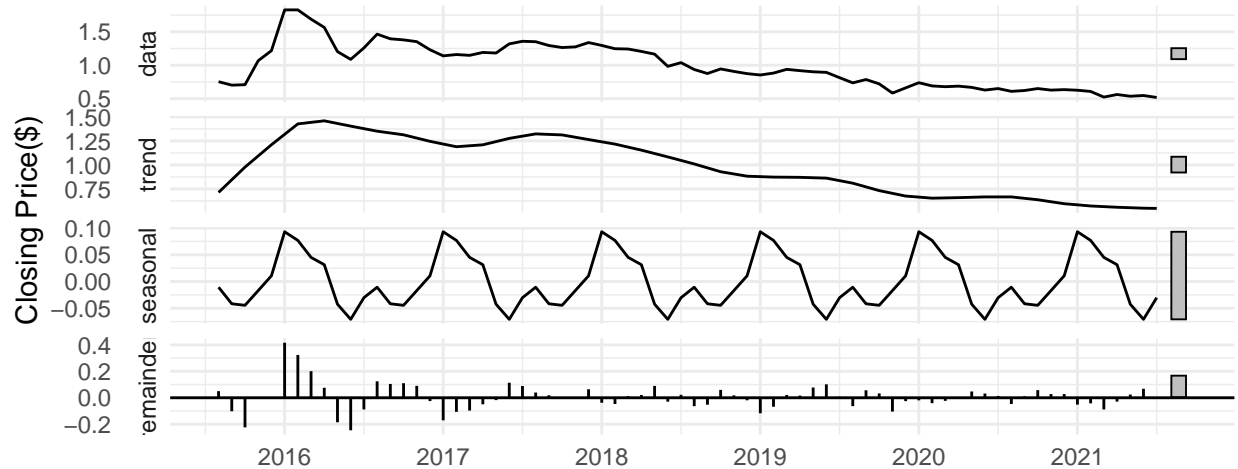
- Trend: There is decreasing trend.

- Seasonality: We see fixed time periods with similar 'visual shapes' on fixed periods within time. This is clear pattern of seasonality.

- Cycles: We see cycles because we have repeating fluctuations for almost every 2 years in trend component.

- Remainder: It is random, hence our seasonal and trend analysis can be considered reliable.

Out-of-graph events

XRP experienced a significant peak in early 2018 due to a surge in cryptocurrency market interest, followed by a sharp decline as market enthusiasm waned and regulatory scrutiny increased.
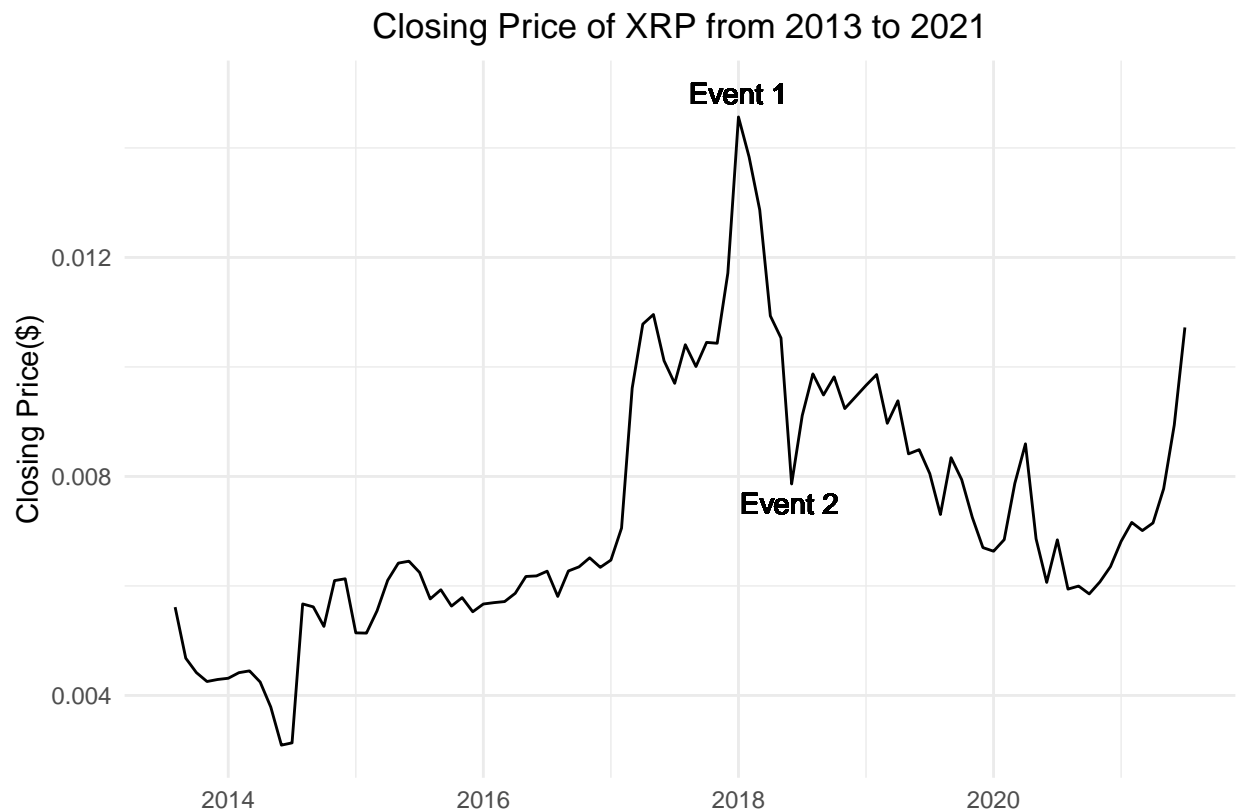
Closing Price of XRP from 2013 to 2021

Event 1 - The end of 2017 and the beginning of 2018 saw a massive surge in the prices of many cryptocurrencies, including XRP. This was partly due to a general bull market in cryptocurrencies, driven by speculative trading and a surge in interest from retail investors. Ripple, the company behind XRP, announced a series of partnerships with financial institutions for using its payment technology, which increased optimism about XRP's adoption for cross-border payments. The prices of cryptocurrencies like Bitcoin soared to all-time highs, many investors jumped into the market out of fear of missing out on potential gains. This led to increased buying pressure on cryptocurrencies, including XRP. The media extensively covered the dramatic rises in cryptocurrency values, which attracted more attention to the space and encouraged new investors to enter the market.

Event 2 - For specific reasons such as regulatory concerns, market sentiment shifts, or technological developments. The decline in XRP's value was consistent with the overall market trend, where many cryptocurrencies saw significant reductions in their value after the speculative bubble of late 2017 and early 2018 burst.

Ethereum's price saw a notable rise in 2016, fueled by technological milestones and the growth of its platform, despite facing volatility from various events.

Closing Price of Ethereum from 2015 to 2021
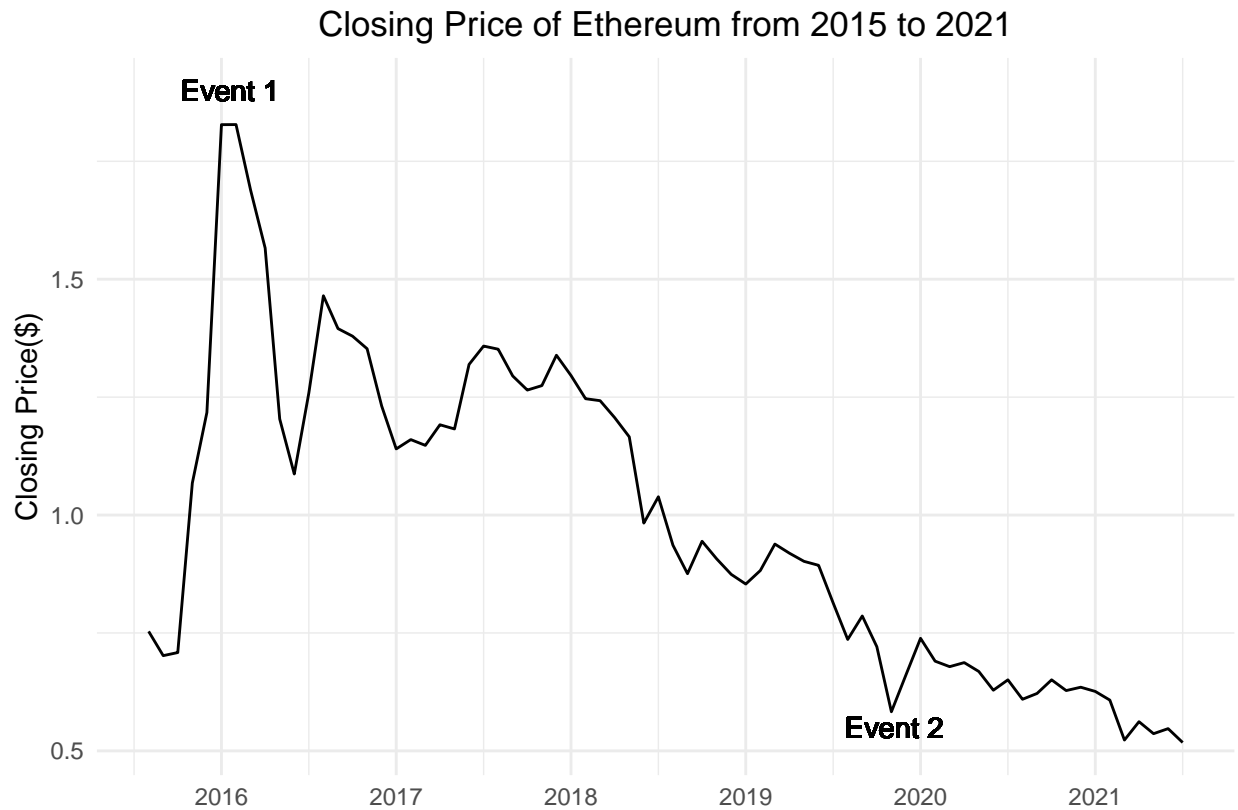
Event 1 - In March, Ethereum released Homestead, the first production version of its software, which was anticipated to provide stability and legitimacy, opening doors for mainstream developers to create decentralized applications on the Ethereum blockchain. Throughout 2016, Ethereum underwent several hard forks to address security issues and improve the network, which demonstrated its flexibility but also raised concerns about its stability as a long-term platform.

Event 2 - Market cycles after a period of rapid growth, a common correction or a sharp decline is noticed as early investors take profits and the market reassesses the value of the asset. Concerns about bans or stringent regulations in major markets like the United States, China, or South Korea could impact Ethereum prices. The rise of other blockchain platforms that compete with Ethereum for developers and user attention impacted negatively.

Wrapped Bitcoin's value generally mirrored Bitcoin's dramatic fluctuations, with a sharp drop before January 2021 after a period of intense growth, influenced by profit-taking and a temporary strengthening of the US dollar.

Closing Price of Wrapped Bitcoin from 2019 to 2021
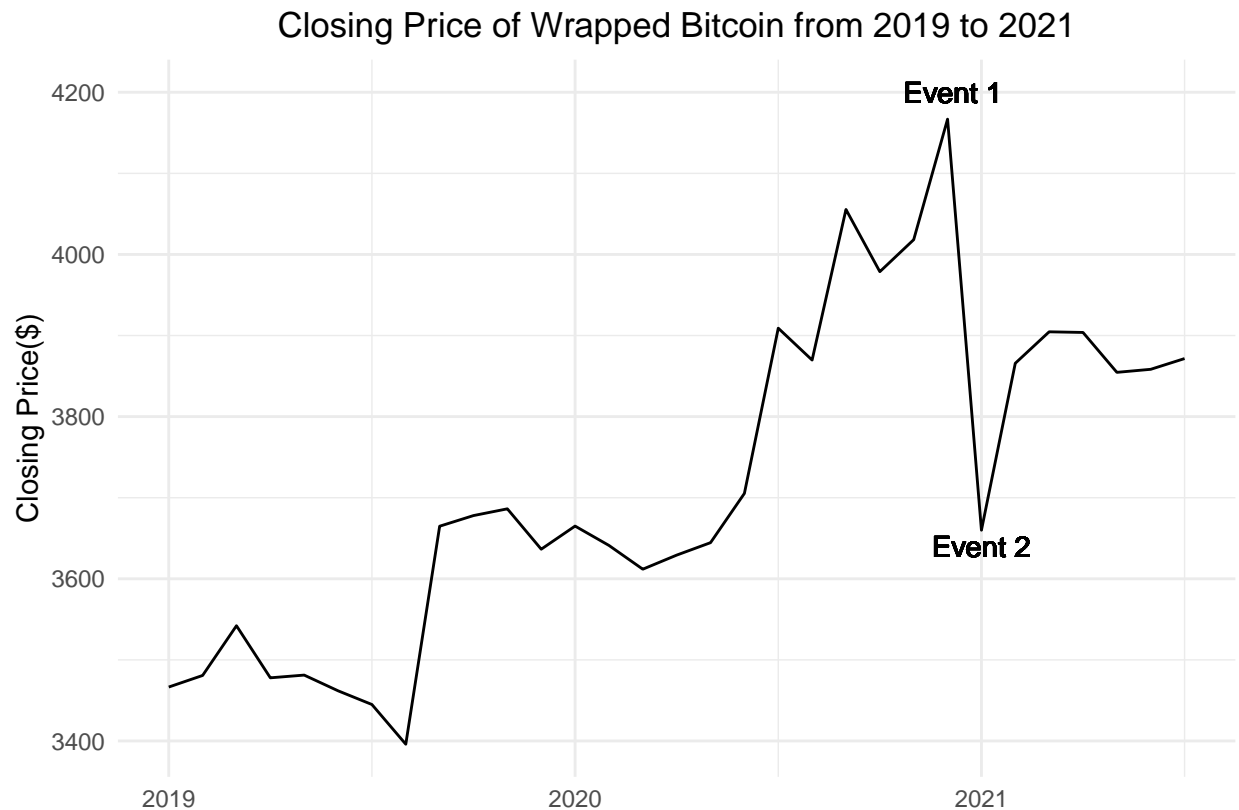
Event 1 - Amidst the economic uncertainty due to the COVID-19 pandemic, there was a significant amount of money printing by central banks around the world, particularly in the US. Investors started to look at Bitcoin as a hedge against potential currency debasement and inflation due to these stimulus measures. 2020 saw increased interest from institutional investors in cryptocurrencies. Traditional finance players like JPMorgan Chase, BlackRock, and others began buying Bitcoin and producing bullish research reports on it. This trend also led to an increased interest in Bitcoin derivatives like Wrapped Bitcoin. Wrapped Bitcoin, as a tokenized version of Bitcoin on the Ethereum network, benefits from the broader trends in the cryptocurrency market, including increased institutional investment and adoption as a hedge against macroeconomic uncertainty.

Event 2 - Wrapped Bitcoin follows due to its nature as a Bitcoin-backed asset on the Ethereum blockchain, reached an all-time high in early January 2021 but then experienced a significant sell-off. This price correction resulted in a loss of 26% of its value, which had a cascading effect on the entire cryptocurrency market, including Wrapped Bitcoin. The rally before this sell-off was driven by the entry of big institutional investors into the market, endorsements from Wall Street figures, and companies like PayPal enabling Bitcoin transactions for its users.

Further analysis of the data set

To analyze the data further, data is compared to take the intersection time period.

```
## -----------------------------------------

## Wrapped Bitcoin - WBTC:

##          Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
## "2019-01-31" "2019-09-09" "2020-04-18" "2020-04-18" "2020-11-26" "2021-07-06"


## -----------------------------------------

## XRP:

##          Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
## "2013-08-05" "2015-07-29" "2017-07-21" "2017-07-21" "2019-07-14" "2021-07-06"


## -----------------------------------------

## Etherium:

##          Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
## "2015-08-08" "2017-01-28" "2018-07-22" "2018-07-22" "2020-01-13" "2021-07-06"


## -----------------------------------------
```

From the summary, the time slot which is common for all currencies is starting from 2019-02-01 up to the last entry of the data set, which is on 2021-07-06.

The points of interest in this case are the changes that happen to the currency. The more are the fluctuations of the selected currency, the more unstable it can be considered. These kind of pattern can be resulted from economic news releases, geopolitical events or unexpected developments.

The percent change of opening prices per day can be calculated. The graph is presented below.
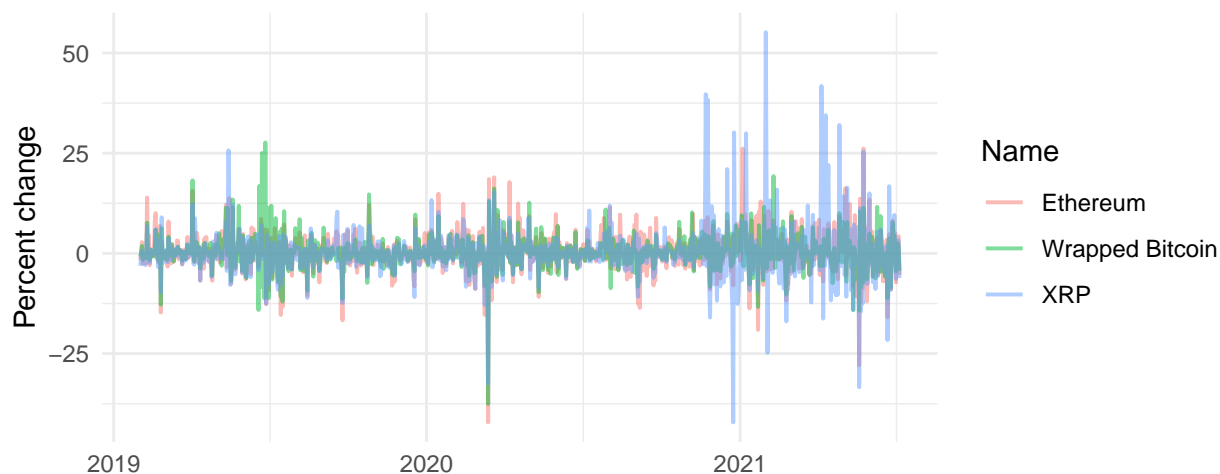
This graph contains too much information and it is not easy to compare the currencies to each other.

Daily percent change density plot

The above graph shows the density plot for the daily change of the opening prices per day. The plot for the wrapped bitcoin is sharper, and has lower variance. All 3 currencies have similar shapes, hence there is nothing that can be told using the density plot.

Standard deviation of Etherium percent change: 5.077022

Standard deviation of Wrapped bitcoin percent change: 4.296524
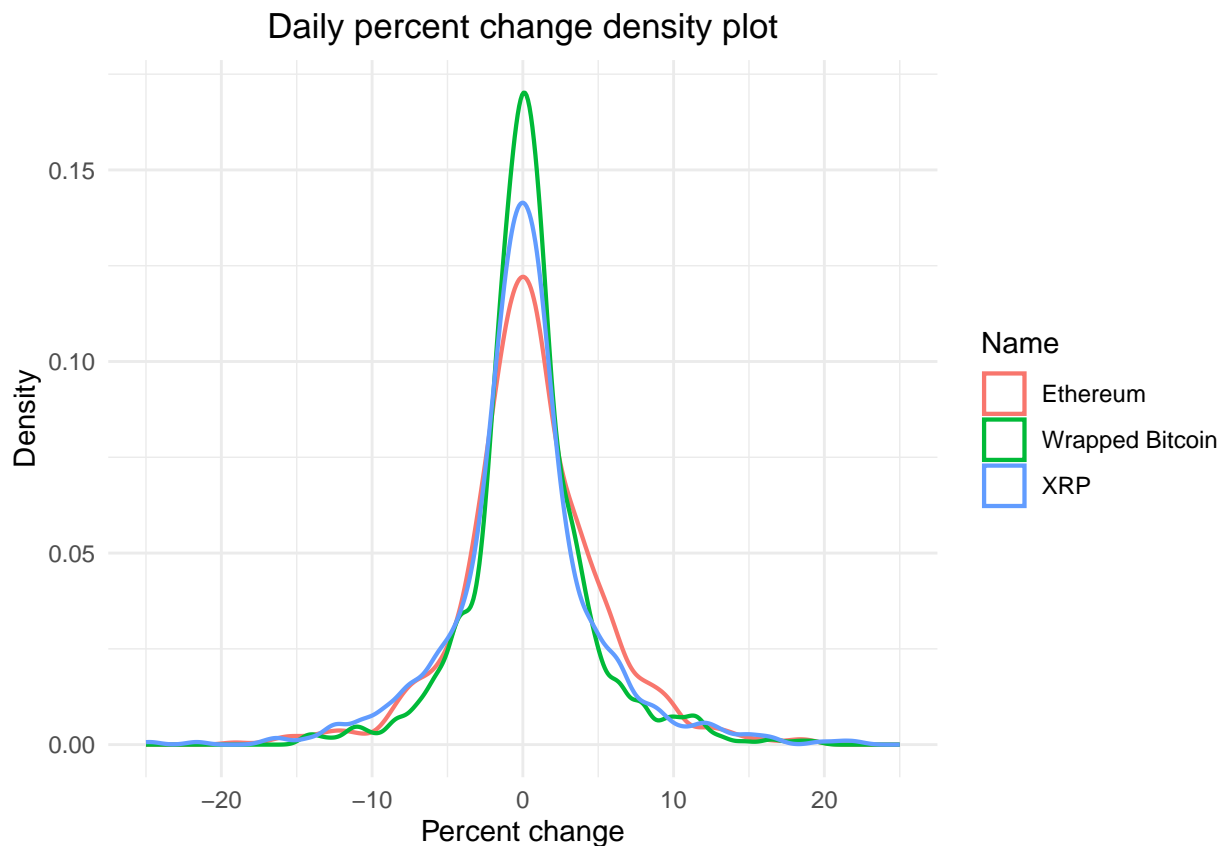
Standard deviation of XRP percent change: 6.451147

Standard deviation values are similar, but XRP has the largest among the given three. Large deviation for the percent change implies that the currency is unstable and can fluctuate more from one day to another.

It might be interesting to see if there is a connection between the given three cryptocurrencies. For that correlation matrix is constructed.

```
##                  Ethereum Wrapped Bitcoin       XRP
## Ethereum        1.0000000       0.7437018 0.6000109
## Wrapped Bitcoin 0.7437018       1.0000000 0.5043311
## XRP             0.6000109       0.5043311 1.0000000
```

Additionally, covariance scatter plot matrix can be constructed to better visualize the values from the covariance matrix.

Scatter plot matrix for the selected cryptocurrencies
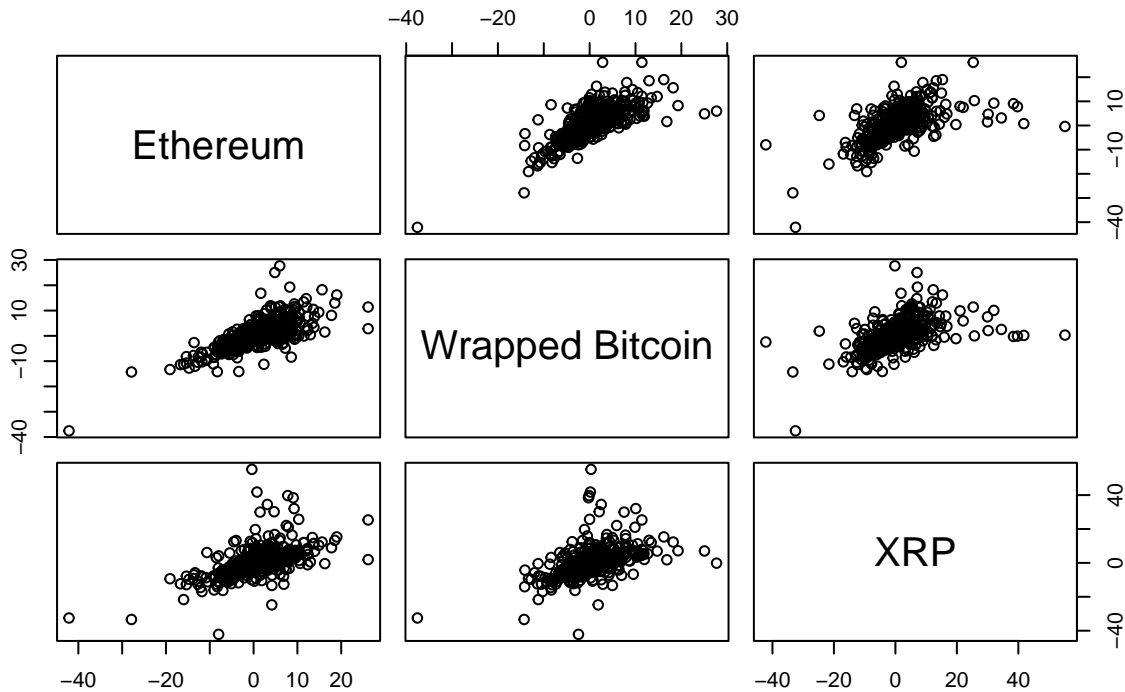
By looking at the given data, we can assume there is no correlation between XRP and the other two currencies, however there is a significant correlation between Etherium and Wrapped Bitcoin. Since wrapped currencies are closely related to their non-wrapped version, relationship between Etherium and Bitcoin isn't surpsrising, but we showed it once again using the correlation. The same can be seen from the scatter plot, as the covariance if Wrapped Bitcoin and XRP are closer to a circle, while Ethereum and Wrapped Bitcoin have a shape closer to an ellipsoid. This shows the relationship between those two currencies one again.

Hotel dataset

Analysis over time

Analysis of average rating and overall customers over time

Average Rating Over Months
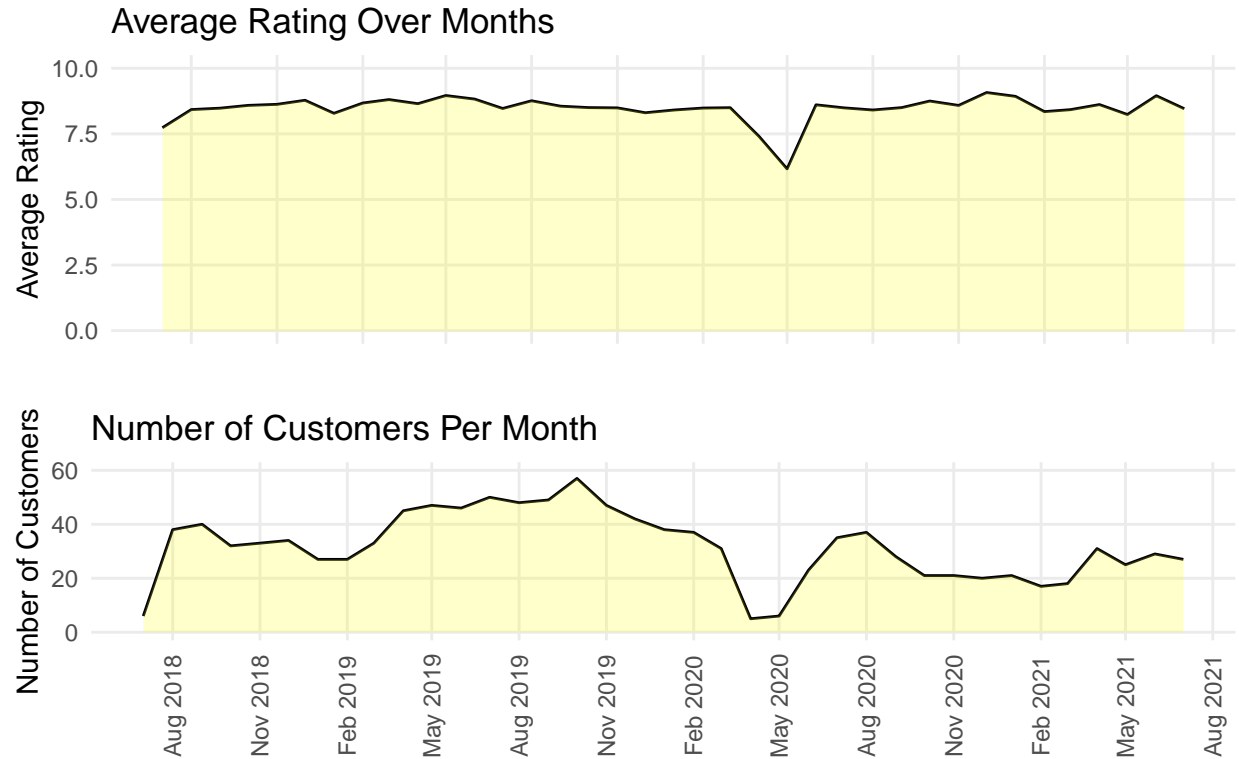
Average Rating Over Months

The plot shows how the average rating changed over time. Since the data entries are not periodic, it can't be treated as time series. However, data per month can be averaged and displayed. The time frame is around 2 years, starting from July 2018 and running up to July 2021. The average rating stays more or less same, around 8. The only visible change in average rating is around May 2020, where the rating decreased to be around 6.5.

Number of Custromers Per Month

To further analyze what happened over time, the number of customers can give an insight. The one regions that catches eye, is the decrease in the number of customers starting from March 2020 until June 2020. This might be caused by the COVID pandemic, when people didn't have the opportunity to travel a lot. The number of customers seems to go up around August, since people are going to holidays and booking hotels at that period of time. The highest peak is around October 2019, which doesn't have any specific reason, since there are no holidays during that time of the year. In comparison with the previous graph, the average rating decreased around the same time when the number of customers decreased. This could have affected the average rating much more, since with less customers, each customer's rating has bigger influence on the average.

Ratings analysis for 5 hotels that have been reviewed the most

Ratings provide a broad understanding of the overall user experience across hotels, so we begin by selecting only the top five hotels that have received the most reviews. This approach aims to enhance the reliability of our findings by minimizing subjectivity, specifically avoiding selection bias.

Ratings analysis for 5 most frequently reviewed hotels
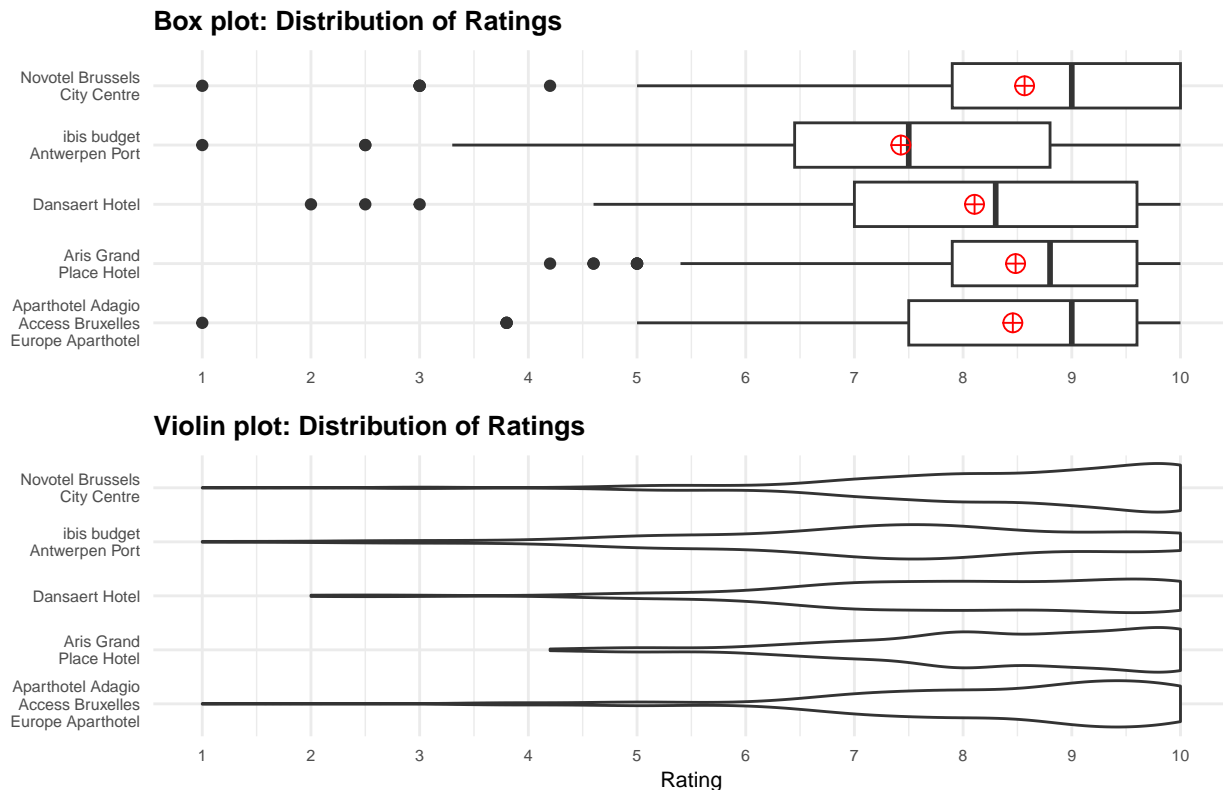
The red sign on each boxplot shows the mean of the distribution of average ratings for each hotel

- For Novotel Brussels City Centre the mean(red circle) is quite smaller than the median(Q2 on the boxplot) meaning that there are relatively few instances with extremely low ratings, pulling the mean downwards. This claim is supported with outliers with low ratings. The long left whisker suggests that there is a notable spread of ratings towards the lower end. Even though the majority of ratings might be high, the presence of the long left whisker implies that there are instances with significantly lower ratings, contributing to the negatively skewed distribution. The box spanning from 7.8 to 10 indicates that the central 50% of ratings is concentrated in this range. This suggests that the majority of ratings fall within a relatively high range, emphasizing the overall positive sentiment of most reviewers. The symmetry in the central 50% of the data suggests that, despite the presence of outliers and a negatively skewed distribution, the majority of ratings within this range are relatively consistent. It could imply that, for the majority of experiences, reviewers tend to provide ratings within high range. Moreover, according to our data 50% of reviewers were from 0 to 9 while the other 50% is fully for 9 to 10.

- For ibis budget Antwerpen Port a mean that is almost equal to the median, which suggests a relatively balanced distribution. The data does not show a strong skewness as the above one, which is why we have whiskers on both sides. Left whisker being longer means that we still have slightly left skewed distribution, with some low ratings. In this case he box spans from 6.5 to 8.8. It indicates that the central 50% of the data is concentrated in this range. This suggests that most ratings fall within this intermediate range. So the previous hotel had better performance.

- For Danaset Hotel the box has larger range, meaning greater variance of central 50% of the ratings. In this case we see slightly left distributed data according to the arrangement of mean, median and whiskers. Central 50% ratings range is about [7, 9.6].

- Aris Grand Place hotel has the most consistent ratings since the outliers are not as extreme as in other cases, it has left skewed distribution, with 50% central ratings ranging from 7.8 to 9.6.

- Finally, Aparthotel Adagio Access Bruxelles Europe Aparthotel has again large variability of ratings with some extremely low ratings but the overall picture is the same(left skewed). A much larger left part of the box indicates that there is more variability, or a larger spread, in the lower ratings. This suggests that there is more diversity in lower customer satisfaction scores.

In order to provide a more user-friendly representation of the analyzed pattern, we use violin plots for each hotel. Violin plots of hotel ratings show the mirrored density, offering a clearer depiction of the spread and concentration of ratings visually, rather than through deep analysis. The wider section of violin plot indicates higher data density, meaning more data points fall within that range.

Using boxplots we can certainly say that hotel 1, 4 and 5 perform quite good, but we cannot be sure which one is better. Violin plot clearly shows that the first hotel has had the most occurrence "10" ratings and the pattern is quite stable(density higher towards higher ratings). We clearly see the skewedness, too.

17

Identifying rating distribution across nationalities

One of the reasons for analyzing this type of data can be to identify target group of these hotels. For simplicity, we take most popular hotels of our data(most popular refers to the number of reviews that the hotel got). We set the popularity threshold to be 80, i.e. any hotel that got at least 80 reviews should have something worth exploring.

To understand what is our target group we look at the nationalities of the reviewers. Hence, from review data of most frequently reviewed hotels, we extract only those reviews where the reviewer has one of the most occurring nationalities in our dataset. We define most occurring nationalities term as nationalities of people who are the most likely to leave some review, i.e. which have the highest occurrence in our data(top 10 in this case).

To get overall picture of what's happening with reviews of reviewers: - who have most popular nationalities, and - who comment on most popular hotels:

Comparison of Ratings by Nationality
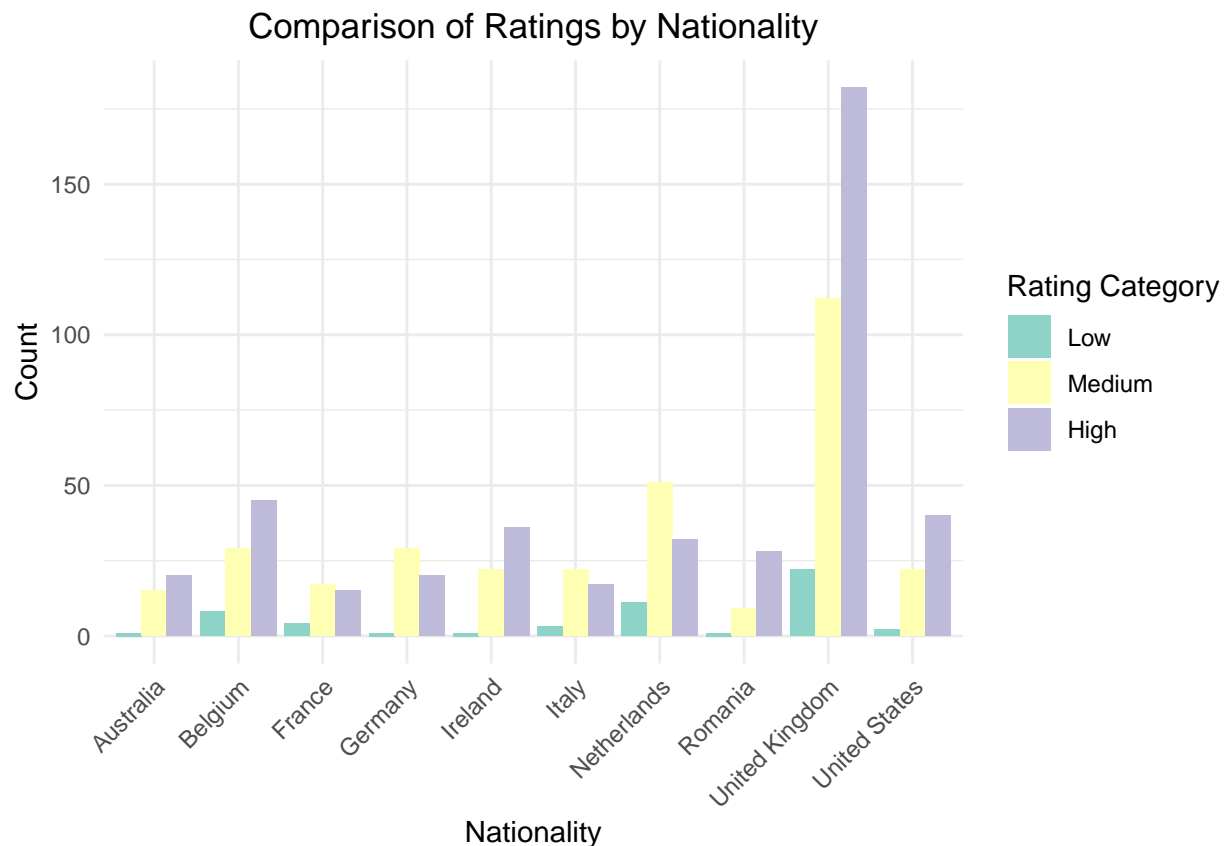
This visualization suggests that: a) People are more likely to leave a neutral or good review, rather than bad. This is because from all depicted nationalities, bars for medium and high ratings are higher (purple and yellow).

  b) We see some obvious outstanding pattern in UK. It has highest bars for each rating category. This means that most of the people reviewing are from UK. This insight can be helpful, since it makes UK people target customers, hence hotels can work on understanding some specifics about the needs of people from this group.

The data is small and doesn't have lots of predictors for diving deeper into the problems described in b), but we can have some understanding about target customers of popular hotels by looking at the tags of their reviews. This way we will understand what's the main purpose and needs of the stay for this group.

To do that we choose to look at the reviewers data from UK only, who commented on popular hotels. The approach is to compare the tags occurrence with ratings to understand: a) whether mentioned factor was a source of satisfaction or dissatisfaction, and b) to know this customer segment better:

Most Occurring Words by Rating Category for UK

Tag	Low	Medium	High
Superior Double Room		25	49
Submitted via mobile	18	73	126
Studio (2 Adults)			16
Stayed 3 nights		16	27
Stayed 2 nights		27	43
Stayed 1 night	12	61	95
Standard Double Room			11
Standard Double or Twin Room		14	25
Solo traveller		30	44
People with friends			13
Leisure trip	17	81	138
Group		11	22
Family with young children		21	30
Couple		48	73
Comfort Double or Twin Room			28
Business trip		27	38

Rating Category

Count — 100, 50

This visualization shows that:

1) Reviews of the people who gave high rating to the mentioned hotels, had 'Submitted via mobile' tag in their taglist(126 overall). This shows that having mobile booking option is a plus since it is convenient feature for customers. Also, convenient option mobile booking can be one of the reasons of reviewer's high rating.

2) Most of the customers(138) come for leisure. It makes sense to focus on that aspect of the hotel.

3) We have some people who come for one night stay who are quite satisfied(61 medium, 95 high) with this hotels.

4) Compared to other factors, whether or not the customer(s) stayed in standard double room or not doesn't play huge role on high rating, since only 11 high ratings were taged with this tag.

5) Even though, leisure trip, 1 night stay, and option for mobile submission play big role on high rating, they are also main reasons for low rating.

6) When giving high rating, people consider a range of factors, but for low rating it is enough to encounter one or slightly more unsatisfactory things, hence the higher the rating, the less empty tiles on heatmap.
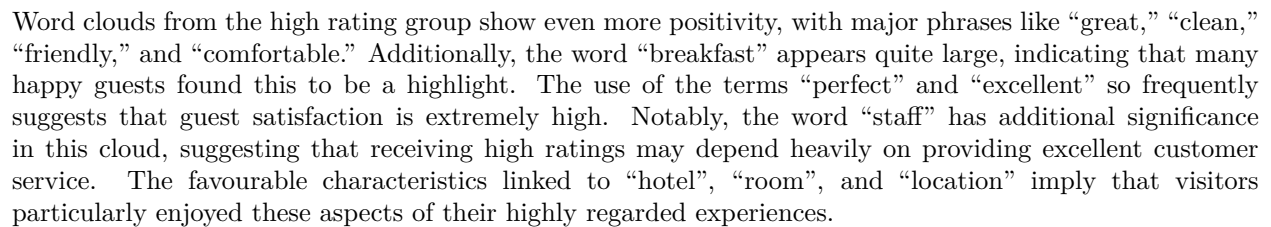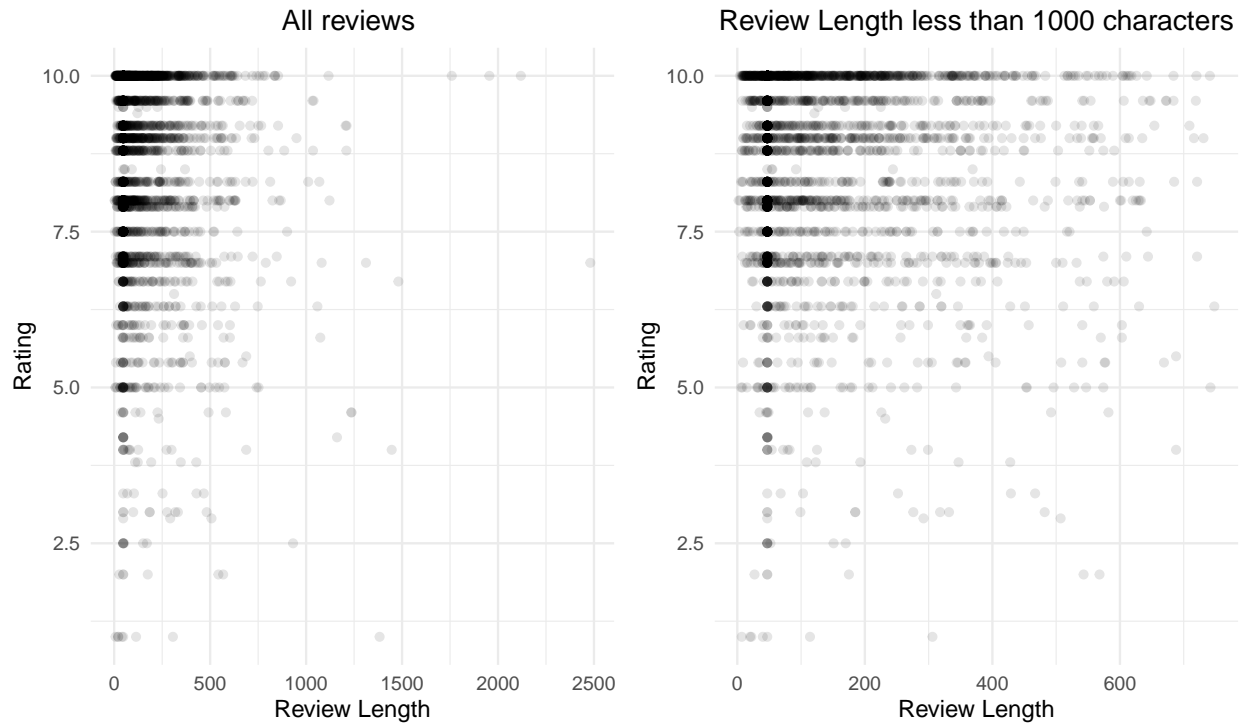
User review analysis

The words "room," "hotel," "staff," and "location" predominate in the word cloud for the low rating group, but words like "dirty," "bad," "small," and "noisy" are also heavily featured. We can make implications that visitors frequently complained about the rooms' size, cleanliness, and noise levels, most probably they complained also about staff performance and other important factors of the hotel.Nevertheless, some positive aspects, such as the "good location" and the "friendly" staff, are still evident despite the overall negative reviews, suggesting that there were some positive aspects even among the lower-rated experiences.

Positive descriptors like "clean", "good", "comfortable", and "friendly" have replaced prominent features like "room", "staff", "location", and "hotel" in the word cloud for the medium rating group. This implies that overall visitor satisfaction was high, with praise for cleanliness and comfort. However, negative words such as "bad", "noise", "noisy" are also in the most used ones with less occurrence, indicating that there were also some dissatisfactions. Words like "breakfast" and "bed" may indicate that the guests found these features satisfactory, which helped the establishment earn a mediocre rating.

Word clouds from the high rating group show even more positivity, with major phrases like "great," "clean," "friendly," and "comfortable." Additionally, the word "breakfast" appears quite large, indicating that many happy guests found this to be a highlight. The use of the terms "perfect" and "excellent" so frequently suggests that guest satisfaction is extremely high. Notably, the word "staff" has additional significance in this cloud, suggesting that receiving high ratings may depend heavily on providing excellent customer service. The favourable characteristics linked to "hotel", "room", and "location" imply that visitors particularly enjoyed these aspects of their highly regarded experiences.

Rating vs Review Length

The scatter plots represent the correlation between the length of a review and the rating given. The plot including all review lengths is not convenient for detailed analyses as there are some outliers, so we generated another plot, which filters reviews to those with less than 1000 characters. There is a dense clustering of reviews at the top of the rating scale, suggesting that many reviews are positive, regardless of length. There doesn't appear to be a strong, consistent pattern indicating that length correlates with higher or lower ratings. We also notice a vertical line at x = 47 (it stands for the sentence "There are no comments available for this review" containing 47 characters. So it does not mean that for length = 47 there is concentration of more ratings. Yet, it's noticeable that there are very few long reviews with low ratings. It might indicate that customers who spend more time writing a review tend to give higher ratings, or perhaps customers with negative experiences tend to leave shorter reviews.