

**American International University- Bangladesh**

**INTRODUCTION TO DATA SCIENCE**

**Finished Project Report**

**Spring 2022-2023**

**Project Title: Crypto and The Block Chain Patent Analysis.**

**Section: B**

|  |  |
| --- | --- |
| Student Name | Student Id |
| Rasel Mahmud | **20-43867-2** |
| Sadia Khanom | **20-44005-2** |
| Tazin Alam | **20-43848-2** |
| A.M. MUFTARIN BILLAH | **19-40813-2** |

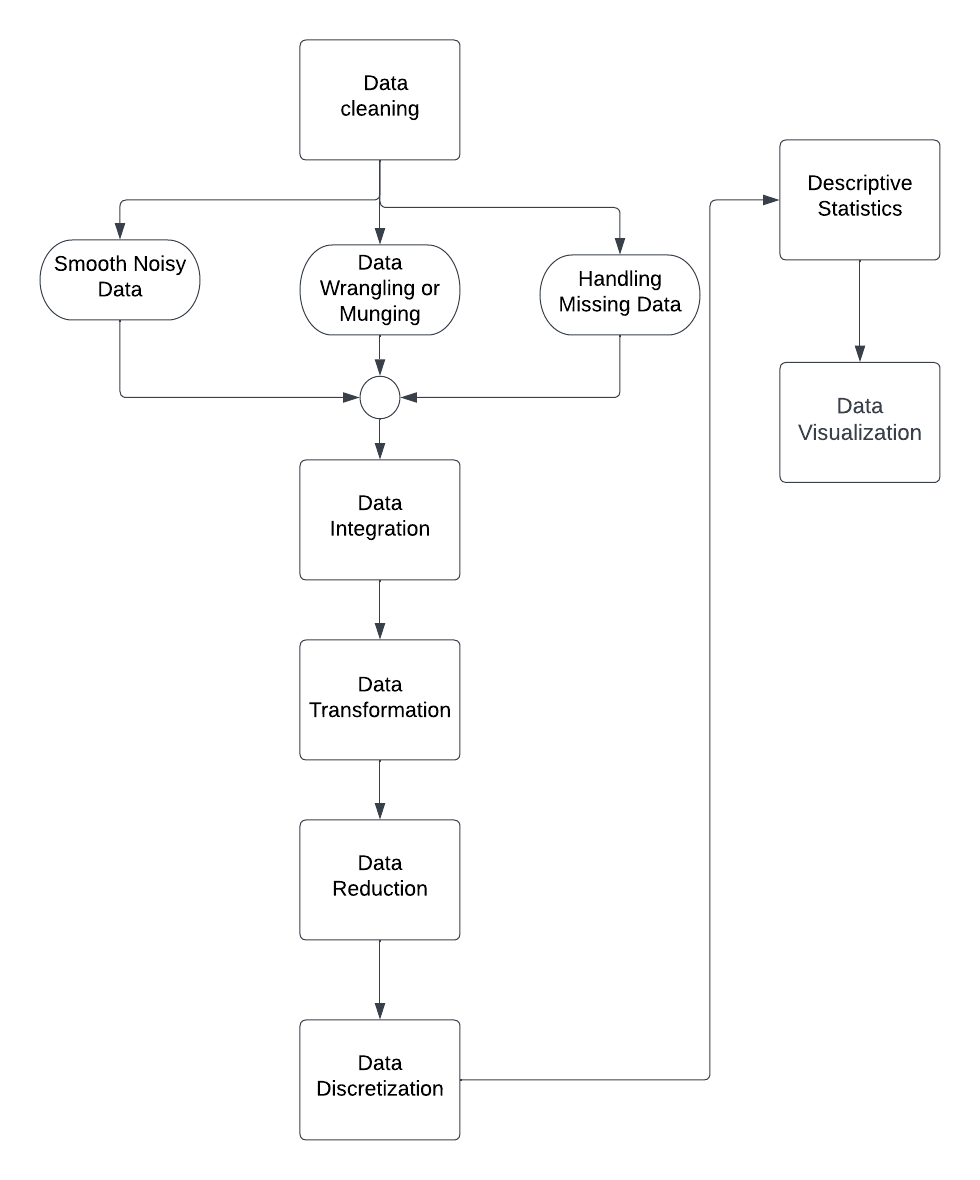
**Project Overview:**

For this project, we have been assigned to scrap data from webpages, perform preprocessing techniques on them, describe them in the light of descriptive statistics and visualize them using R language.

In our project firstly, we chose crypto data of the top 100 crypto in the cryptocurrency market or the digital currency market. We collected crypto data from famous cryptocurrency exchange market website <https://www.coingecko.com/>. We Collected the data in 27th April. After Gathering Data from the cryptocurrency exchange market website we do some calculations to predict in the next day which crypto price from the top 100 crypto may rise and in the next day in which crypto the investor may interested most. After that we collect the Blockchain patent ownership information from a website. Patent information can be significant in the crypto market for a few reasons. First, it can provide insights into the development and innovation of new technologies and products within the industry. Second, patent information can also provide clues about which companies are investing in crypto-related research and development. This information can be useful for investors looking to make informed decisions about which companies to invest in. Finally, patent information can also be relevant in legal disputes within the industry. In the event of a patent infringement lawsuit, the information contained in the patents can be used as evidence to support or challenge claims of ownership and infringement. We collected the Patent information from the website <https://harrityllp.com/titans-of-technology-blockchain-the-top-companies-in-blockchain-patents-2021/>. Data Transformation, Data Reduction, and Data Discretization. We did data pre-processing where it was needed. In Descriptive analysis, we described our data with the help of descriptive methods. In the descriptive analysis, we describe our data in some manner and present it in a meaningful way so that it can be easily understood. To describe a comparison between different things we did the Mean, Median, Mode, Range, Variance, Quartile & Percentile. Lastly, we did data visualization to see and understand as visualizations can more effectively allow the reader to digest information. Graphics can allow users to deliver insights in a much easier fashion than describing through text and can also have a greater impact. Here we tried to visualize almost every aspect of comparison & relation.

**Project Solution Design:**

We initially gathered our player lists and performance information for Crypto and and Patent Content from several websites in order to prepare the dataset for data analysis. We then recorded the information in a CSV file. The data pre-processing is then done. Data cleaning is the process of inspecting a raw dataset to find and eliminate errors, duplication, and superfluous data. The table had some missing data, which we replaced with N/A and then filled up with the median. Then we tried to manage every item of noisy data that was in the dataset. After performing data cleaning, measures for data integration, data transformation, data reduction, and data discretization were taken to further clean the data set. We concentrated on using descriptive statistics to rationally simplify our enormous volumes of data after completing the data preprocessing. Moreover, to sum up, the dataset's approximate data. In our data collection, we used the following metrics: Mean, Median, Mode, Range, Variance, Standard Deviation, Quartiles, Percentiles, and Interquartile Ranges. We used data visualization to present facts and data graphically after finishing the descriptive statistics.



**Data Collection:**

For this project, we start to scrap the data from the website. First, we start to scrap the data from team Coingeko Website for the top 100 crypto market status. In this process, we use a selector gadget to simply select data on a website .

**Getting CryptoData:**

**Graphical user interface, text

Description automatically generated**

**Code:**

|  |
| --- |
| library(rvest)  library(dplyr)  link="https://www.coingecko.com/"  page=read\_html(link)  Coin=page%>%html\_nodes(".font-bold")%>%html\_text()  Coin <- Coin[-c(101,102)]  Price=page%>%html\_nodes(".tw-flex-1 .no-wrap")%>%html\_text()  last\_1\_hour=page%>%html\_nodes(".change1h span")%>%html\_text()  last\_24\_hour=page%>%html\_nodes(".change24h span")%>%html\_text()  last\_7\_days=page%>%html\_nodes(".change7d span")%>%html\_text()  last\_24hours\_volume=page%>%html\_nodes(".lit .no-wrap")%>%html\_text()  Mkt\_Cap=page%>%html\_nodes(".cap-price .no-wrap")%>%html\_text()  #Data Frame1  Crypto1=data.frame(Coin,Price,last\_1\_hour,last\_24\_hour,last\_7\_days,last\_24hours\_volume,Mkt\_Cap,stringsAsFactors = FALSE) |

**Getting Patent Data:**

**A picture containing text

Description automatically generated**

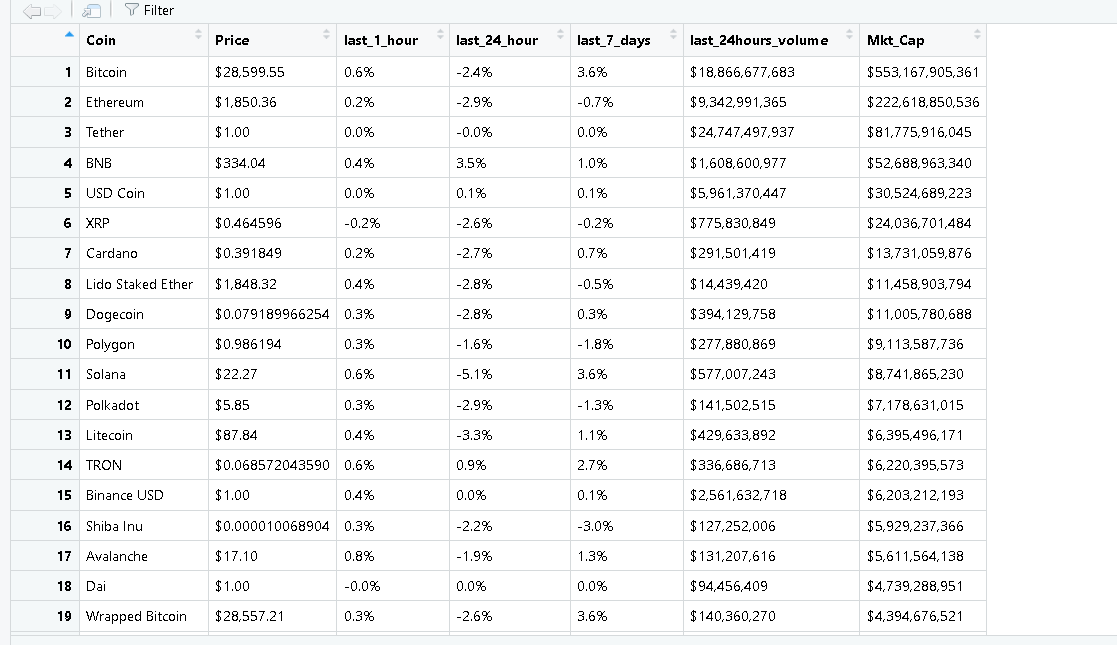
**Text

Description automatically generated with medium confidence**

**Code:**

|  |
| --- |
| > library(rvest)  > library(dplyr)  > response <- read\_html("https://harrityllp.com/titans-of-technology-blockchain-the-top-companies-in-blockchain-patents-2021/")  > html\_text(response)  > tables <- response %>% html\_nodes("table") %>% html\_table()  > table\_one = tables[[1]]  > table\_two = tables[[2]]  > table\_three = tables[[3]]  > table\_four = tables[[4]]  > table\_five = tables[[5]]  > table\_six = tables[[6]]  > table\_seven = tables[[7]]  > Top\_Countries\_in\_Blockchain\_Patents\_2021=data.frame(table\_one,stringsAsFactors = FALSE)  > Top\_Companies\_in\_Worldwide\_Blockchain\_Patents\_PendingApplications=data.frame(table\_two,stringsAsFactors = FALSE)  > Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021=data.frame(table\_three,stringsAsFactors = FALSE)  > Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021=data.frame(table\_four,stringsAsFactors = FALSE)  > Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021=data.frame(table\_five,stringsAsFactors = FALSE)  > Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021=data.frame(table\_six,stringsAsFactors = FALSE)  > IBM\_vs\_Advanced\_New\_Technologies\_Blockchain\_Competitive\_Gap\_Analysis\_2021=data.frame(table\_seven,stringsAsFactors = FALSE)  > write.csv(Top\_Countries\_in\_Blockchain\_Patents\_2021, "Top\_Countries\_in\_Blockchain\_Patents\_2021.csv", row.names = FALSE)  > write.csv(Top\_Companies\_in\_Worldwide\_Blockchain\_Patents\_PendingApplications, "Top\_Companies\_in\_Worldwide\_Blockchain\_Patents\_PendingApplications.csv", row.names = FALSE)  > write.csv(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021, "Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021.csv", row.names = FALSE)  > write.csv(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021, "Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021.csv", row.names = FALSE)  > write.csv(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021, "Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021.csv", row.names = FALSE)  > write.csv(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021, "Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021.csv", row.names = FALSE)  > write.csv(IBM\_vs\_Advanced\_New\_Technologies\_Blockchain\_Competitive\_Gap\_Analysis\_2021, "IBM\_vs\_Advanced\_New\_Technologies\_Blockchain\_Competitive\_Gap\_Analysis\_2021.csv", row.names = FALSE) |

**Out Put of The Data Frames:**

****

**Table

Description automatically generated**

**Table

Description automatically generated**

**Data Pre-processing:**

Now the most important phase of the data analysis starts which is data pre-processing. We are going to use pre-processing techniques on the datasets to prepare a well completed datasets for analysis and visualization.

1. **Data Cleaning**

* **Handling Missing Data:** To handle missing data we first need to search the data set for any value that is not assigned. To do so we write a code that will show us the row which contains the missing value,

Code:

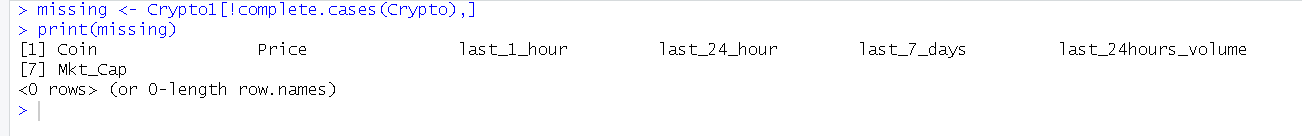
|  |
| --- |
| missing1 <- Top\_Countries\_in\_Blockchain\_Patents\_2021[!complete.cases(Top\_Countries\_in\_Blockchain\_Patents\_2021),]  print(missing)  missing2 <- Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021[!complete.cases(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021),]  print(missing)  missing3 <- Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021[!complete.cases(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021),]  print(missing)  missing4 <- Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021[!complete.cases(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021),]  print(missing)  missing5 <- Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021[!complete.cases(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021),]  print(missing) |

|  |
| --- |
| missing <- Crypto1[!complete.cases(Crypto),]  print(missing) |

**Output:**

 Graphical user interface, text, application, email

Description automatically generated



Here We can see that There is no Missing Data in the Data Frames. So Here We don not need to deal with any missing value and we do not need to worry about the missing values.

* **Smooth Noisy Data:** In the dataset, we can see that some columns contain a mixture of both numerical and character data. Like in the Crypto Data Frame Price, Market Cap, Last 24 Volume Contains the contains extra $ sign and last 1h, last 24h and last 7 day the performance contains % as a parameter. For the betterment of the calculation, we have to remove those noises from the dataset.

Code:  
In the Crypto Data Frame to remove the $ and , in Price, Market Cap, Last 24 Volume

|  |
| --- |
| Crypto1$Price <- sub("\\$", "", Crypto1$Price)  Crypto1$Mkt\_Cap <- sub("\\$", "", Crypto1$Mkt\_Cap)  Crypto1$last\_24hours\_volume <- sub("\\$", "", Crypto1$last\_24hours\_volume)  Crypto1$Price <- gsub(",", "", Crypto1$Price)  Crypto1$Mkt\_Cap <- gsub(",", "", Crypto1$Mkt\_Cap)  Crypto1$last\_24hours\_volume <- gsub(",", "", Crypto1$last\_24hours\_volume) |

In the Crypto Data Frame to remove the % in last 1h, last 24h and last 7 day.

|  |
| --- |
| Crypto1$last\_1\_hour <- sub("%", "", Crypto1$last\_1\_hour)  Crypto1$last\_24\_hour <- sub("%", "", Crypto1$last\_24\_hour)  Crypto1$last\_7\_days <- sub("%", "", Crypto1$last\_7\_days) |

**Out Put:**

**Table

Description automatically generated**

There is no need to Smooth rest of the Data Frames.

* **Data Munging:** The dataset does not require munging because all the data are within the same range.

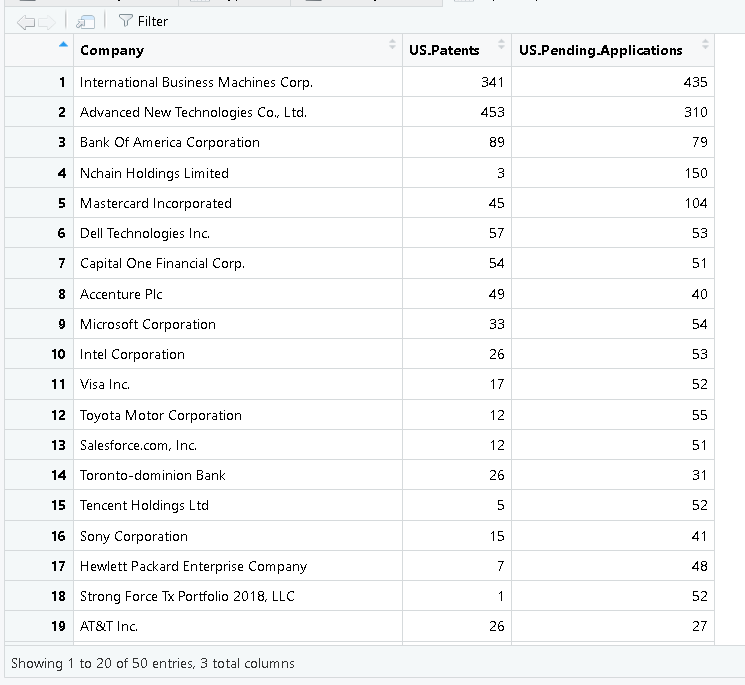
1. **Data Reduction:**

In our patent holder countries data frames, we want to focus only the top 50 blockchain patent institute/company from every country so we can Reduce the Number of Rows from the data frames.

Code:

|  |
| --- |
| Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021 <- head(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021, 50)  Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021 <- head(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021, 50)  Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021 <- head(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021, 50)  Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021 <- head(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021, 50) |

**A screenshot of a computer

Description automatically generated with low confidence**

1. **Data Integration:**

For the purpose of better analysis, we need to add two Extra Column in Each of the patent holder crypto country name “Country Origin” and “Grand Total Number of Patent”. And in the Crypto Data Frame we will add a Extra Column Name “Track of the Price Percentage” which will hold the summation of price up down percentage from last 7 days to till today.

Patent

Table

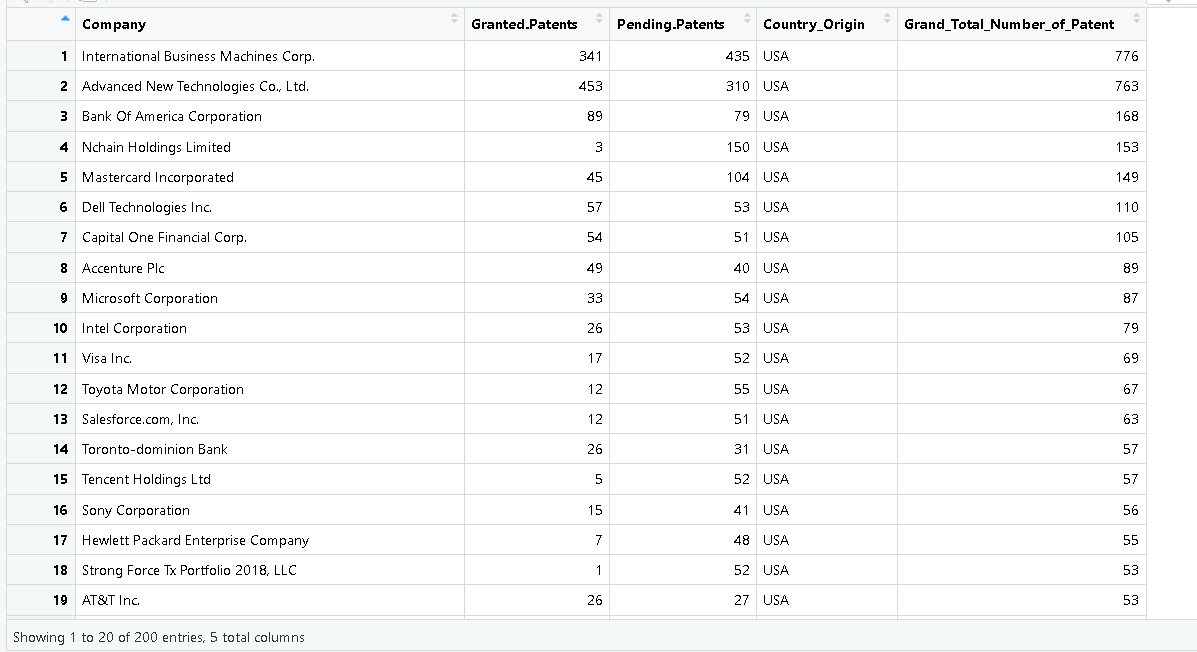
Description automatically generated

Table

Description automatically generated

For the purpose of better analysis, we need to integrate four patent data frames into one complete dataset.

|  |
| --- |
| > data <- rbind(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021,Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021,Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021,Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021)  > View(data) |



A new Column Scale the number of grand total patent in which Grand Total Patent less than 100 or Equal is Scale as Low, Grand Total Patent less than or equal 500 and More than 100 is categorized as Medium Scale, Grand Total Patent less than or equal 1000 and More than 500 is categorized as Medium Large Scale, Grand Total Patent less than or equal 1500 and More than 1000 is categorized as Large Scale and age greater than 1500s is categorized as Excellent Scale.

|  |
| --- |
| data$Grand\_Total\_Number\_of\_Patent <- as.numeric(data$Grand\_Total\_Number\_of\_Patent)  data$Patent\_Scale <- case\_when(  data$Grand\_Total\_Number\_of\_Patent <= 100 ~ "Low",  data$Grand\_Total\_Number\_of\_Patent <= 500 ~ "Medium",  data$Grand\_Total\_Number\_of\_Patent <= 1500 ~ "Large",  data$Grand\_Total\_Number\_of\_Patent > 1500 ~ "Excellent"  ) |

For a better understanding of the crypto, we integrate a new column named Performance\_Progress, which is the sum of the last 1h, last 24h and last 7 days performance.

A new two column named Performance\_Progress which is the sum of last 1h, last 24h and last 7 days performance and another One is Growth added in the Crypto1 Data Frame for the better Analysis in Future.

|  |
| --- |
| Crypto1$Performance\_Progress <-  Crypto1$last\_1\_hour +  Crypto1$last\_24\_hour+Crypto1$last\_7\_days  Crypto1$Performance\_Progress <- as.numeric(Crypto1$Performance\_Progress)  Crypto1$Growth <- case\_when(  Crypto1$Performance\_Progress <0 ~ "Negative",  Crypto1$Performance\_Progress ==0 ~ "No Change",  Crypto1$Performance\_Progress >3 ~ "Large Growth",  Crypto1$Performance\_Progress >0 ~ "Positive"  ) |

Out Put:

Table

Description automatically generated

1. **Data Transformation**

Here we need to convert some of the Data Frames columns values numeric.

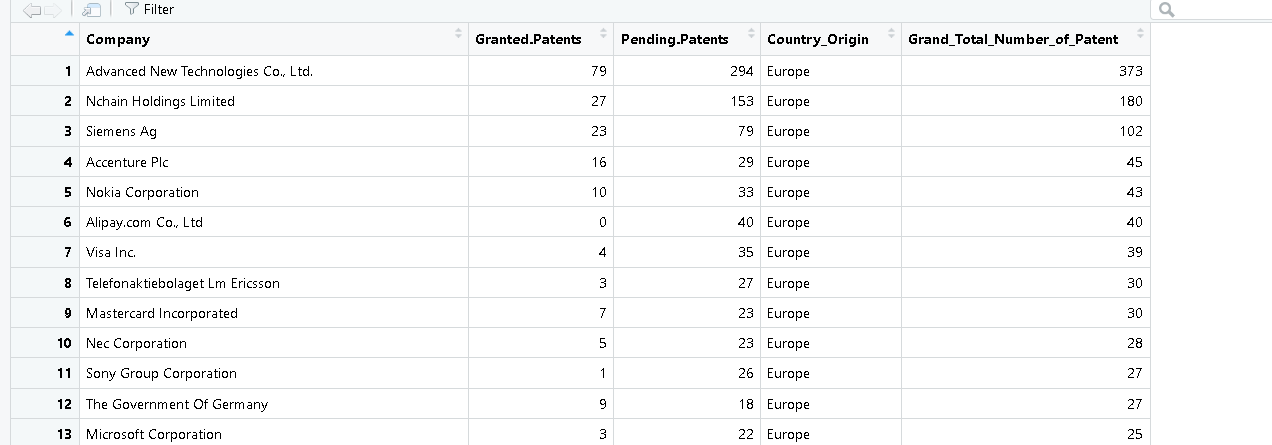
Code:

|  |
| --- |
| #usa  Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021$US.Patents <- as.numeric(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021$US.Patents)  Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021$US.Pending.Applications <- as.numeric(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021$US.Pending.Applications)  #china  Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021$Chinese.Patents <- as.numeric(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021$Chinese.Patents)  Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021$Chinese.Pending.Applications <- as.numeric(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021$Chinese.Pending.Applications)  #korean  Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021$Korean.Patents <- as.numeric(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021$Korean.Patents)  Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021$Korean.Pending.Applications <- as.numeric(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021$Korean.Pending.Applications)  #Europian  Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021$EPO.Patents <- as.numeric(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021$EPO.Patents)  Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021$EPO.Pending.Applications <- as.numeric(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021$EPO.Pending.Applications) |

We need to change some of the column name in the data frames of the Patent so that we can do the integration between the Data Frames.

Code:

|  |
| --- |
| colnames(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021)[colnames(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021)=="US.Patents"] <- "Granted.Patents"  colnames(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021)[colnames(Top\_Companies\_in\_US\_Blockchain\_Patents\_Pending\_Applications\_2021)=="US.Pending.Applications"] <- "Pending.Patents"  colnames(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021)[colnames(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021)=="Chinese.Patents"] <- "Granted.Patents"  colnames(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021)[colnames(Top\_Companies\_in\_Chinese\_Blockchain\_Patents\_Pending\_Applications\_2021)=="Chinese.Pending.Applications"] <- "Pending.Patents"  colnames(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021)[colnames(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021)=="Korean.Patents"] <- "Granted.Patents"  colnames(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021)[colnames(Top\_Companies\_Korean\_Blockchain\_Patents\_Pending\_Applications\_2021)=="Korean.Pending.Applications"] <- "Pending.Patents"  colnames(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021)[colnames(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021)=="EPO.Patents"] <- "Granted.Patents"  colnames(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021)[colnames(Top\_Companies\_in\_European\_Blockchain\_Patents\_Pending\_Applications\_2021)=="EPO.Pending.Applications"] <- "Pending.Patents" |



We have converted the some of Crypto DataFrame column value as numeric for better Analysis.

|  |
| --- |
| Crypto1$last\_1\_hour <- as.character(Crypto1$last\_1\_hour)  Crypto1$last\_24\_hour <- as.numeric(Crypto1$last\_24\_hour)  Crypto1$last\_7\_days <- as.numeric(Crypto1$last\_7\_days)  Crypto1$Price <- as.numeric(Crypto1$Price)  Crypto1$Mkt\_Cap <- as.numeric(Crypto1$Mkt\_Cap)  Crypto1$last\_24hours\_volume <- as.numeric(Crypto1$last\_24hours\_volume) |

we need to transform some variables for better analysis of the dataset.

|  |
| --- |
| Crypto1$Coin <- factor(Crypto1$Coin, ordered = TRUE)  Crypto1$Growth <- factor(Crypto1$Growth,  levels =c(1,2,3,4),labels=c("Large Growth","Positive","No Change","Negative"))  data$Country\_Origin <- factor(data$Country\_Origin, ordered = TRUE)  data$Patent\_Scale <- factor(data$Patent\_Scale,  levels =c(1,2,3,4),labels=c("Excellent","Large","Medium","Low")) |

**5.Data Discretization:**

No discretization is needed for this dataset as it is already in a better shape. So we skip this process and move on to descriptive statistics.

**Descriptive Statistics:**

Now, we are going to compute various descriptive statistics parameters for our dataset.

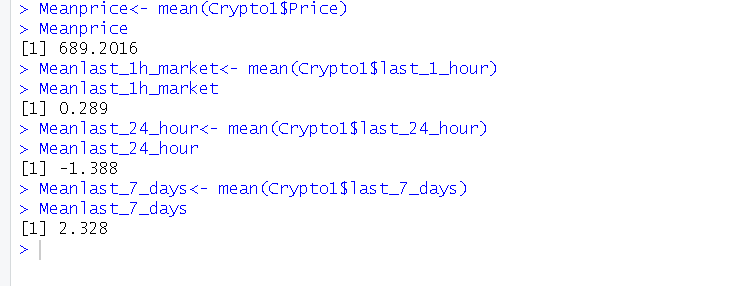
Firstly, let’s try to inspect the central tendency for the various variables of our dataset.

* **MEAN:**

Mean of all top 100 crypto last\_1h, last\_24h and last\_7day market status in Crypto Data Frame.

Code:

|  |
| --- |
| Meanprice<- mean(Crypto1$Price)  Meanprice  Meanlast\_1h\_market<- mean(Crypto1$last\_1\_hour)  Meanlast\_1h\_market  Meanlast\_24\_hour<- mean(Crypto1$last\_24\_hour)  Meanlast\_24\_hour  Meanlast\_7\_days<- mean(Crypto1$last\_7\_days)  Meanlast\_7\_days |



* **MEDIAN:**

Now we calculate the median for the last\_24 Market Cap and Volume of the Crypto

Code:

|  |
| --- |
| medianone=sort(Crypto1$Mkt\_Cap)  medianoneres=median(medianone)  medianoneres  mediantwo=sort(Crypto1$last\_24hours\_volume)  mediantwores=median(mediantwo)  mediantwores |

OutPut

Text

Description automatically generated

* **MODE:**

As the mode doesn’t have a built-in function, we first implement the function.

Code:

mode <- function(x){

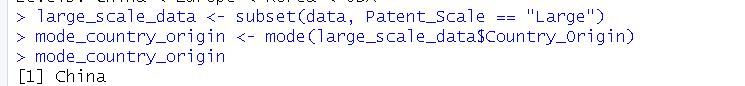
unique\_values <- unique(x)

table <- tabulate(match(x, unique\_values))

unique\_values[table == max(table)]

}

|  |
| --- |
| large\_scale\_data <- subset(data, Patent\_Scale == "Large")  mode\_country\_origin <- mode(large\_scale\_data$Country\_Origin)  mode\_country\_origin  excellent\_scale\_data <- subset(data, Patent\_Scale == "Excellent")  mode\_country\_origin <- mode(excellent\_scale\_data $Country\_Origin)  mode\_country\_origin  medium\_scale\_data <- subset(data, Patent\_Scale == "Medium")  mode\_country\_origin <- mode(medium\_scale\_data $Country\_Origin)  mode\_country\_origin  small\_scale\_data <- subset(data, Patent\_Scale == " Low ")  mode\_country\_origin <- mode(small\_scale\_data $Country\_Origin)  mode\_country\_origin |



Text

Description automatically generated

Text

Description automatically generated with medium confidence

Text

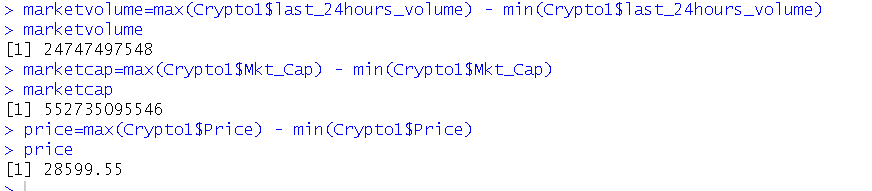
Description automatically generated with low confidence

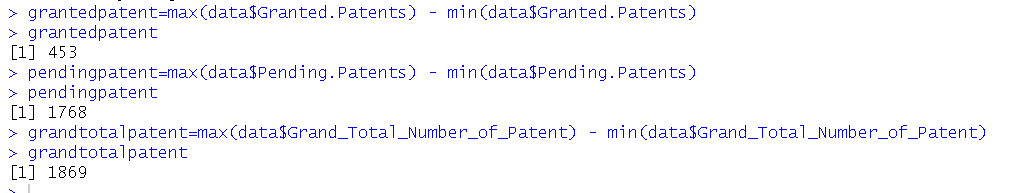
**Range:**

Now we calculate the range of variables.

Code:

|  |
| --- |
| marketvolume=max(Crypto1$last\_24hours\_volume) - min(Crypto1$last\_24hours\_volume)  marketvolume  marketcap=max(Crypto1$Mkt\_Cap) - min(Crypto1$Mkt\_Cap)  marketcap  price=max(Crypto1$Price) - min(Crypto1$Price)  price  grantedpatent=max(data$Granted.Patents) - min(data$Granted.Patents)  grantedpatent  pendingpatent=max(data$Pending.Patents) - min(data$Pending.Patents)  pendingpatent  grandtotalpatent=max(data$Grand\_Total\_Number\_of\_Patent) - min(data$Grand\_Total\_Number\_of\_Patent)  grandtotalpatent |





**Quartile & Percentile:**

Code:

|  |
| --- |
| quantile(Crypto1$last\_1\_hour, prob = c(0.0,0.25,0.50, 0.75 , 0.100))  quantile(Crypto1$last\_24\_hour, prob = c(0.0,0.25,0.50, 0.75 , 0.100))  quantile(Crypto1$last\_7\_days) |

Text

Description automatically generated

**Interquartile Range:**

Code:

IQR(Crypto1$Mkt\_Cap)

A picture containing text

Description automatically generated

**Variance:**

Code:

var(Crypto1$last\_1\_hour)

var(Crypto1$last\_24\_hour)

var(Crypto1$last\_7\_days)

Output:

Graphical user interface, text, application

Description automatically generated

**Standard Deviation:**

Code:

sd(Crypto1$last\_1\_hour)

sd(Crypto1$last\_24\_hour)

sd(Crypto1$last\_7\_days)

Text

Description automatically generated with medium confidence

**Normal Distribution:**

Code:

|  |
| --- |
| x = rnorm(Crypto1$last\_1\_hour, mean = mean(Crypto1$last\_1\_hour), sd=  sd(Crypto1$last\_1\_hour))  hist(x) |

Chart, histogram

Description automatically generated

Code:

|  |
| --- |
| y = rnorm(Crypto1$last\_24\_hour, mean = mean(Crypto1$last\_24\_hour), sd=  sd(Crypto1$last\_24\_hour))  hist(y) |

Chart, histogram

Description automatically generated

|  |
| --- |
| z = rnorm(Crypto1$last\_7\_days, mean = mean(Crypto1$last\_7\_days), sd=  sd(Crypto1$last\_7\_days))  hist(z) |

Chart, histogram

Description automatically generated

**Data Visualization:**

1. Now the most important part of the visualization. We need to see the Overall Week Growth of the top 100 Cryptos.

Code:

|  |
| --- |
| library(ggplot2)  Crypto1$Market\_Cap\_Quartile <- cut(Crypto1$Mkt\_Cap, quantile(Crypto1$Mkt\_Cap, probs = c(0, 0.25, 0.5, 0.75, 1)), include.lowest = TRUE, labels = c("$1M-$1B", "$1B-$10B", "$10B-$100B", "$100B-$1T"))  # Create a bar plot of growth by market cap quartile  ggplot(Crypto1, aes(x=Growth, fill=Market\_Cap\_Quartile)) +  geom\_bar(position = position\_fill(reverse = TRUE)) +  labs(title = "Overall Week Growth Reports by Market Cap Quartile", x = "Performance", y = "Frequency") +  coord\_flip() |

Chart, bar chart

Description automatically generated

This Weekly Report Shows that Most of the coin overall growth remains the same.

2)

Chart, scatter chart

Description automatically generatedThis Scatter Plot Shows the Scatter Plot for Total Market Cap and Last 24 hour volume of the top 100 cryptos in the Coingeko.

The Crypto Which has the Large Market Cap has the Largest Volume in the Last 24 Hour Market Cap. It means the Crypto which has the Largest Market Cap Investor will invest more on these Cryptos as they have a bit lower chance for fall down certainly.

3)

|  |
| --- |
| filtered\_data <- Crypto1 %>%  filter(Growth == "Positive", Mkt\_Cap > 10000000000)  # Create scatter plot of market cap vs. growth for filtered data  ggplot(filtered\_data, aes(x = Coin, y = Mkt\_Cap, fill = Coin)) +  geom\_bar(stat = "identity", position = "dodge") +  scale\_y\_log10() +  labs(title = "Coins with Large Growth and High Market Cap", x = "Coin", y = "Market Cap ($)") +  scale\_fill\_manual(values = rainbow(length(unique(filtered\_data$Coin)))) +  theme\_bw() |

Chart, bar chart

Description automatically generated

This Graph Shows the Top 4 coin in crypto Market With Large Growth.

|  |
| --- |
| filtered\_data <- Crypto1 %>%  filter(Growth == "Large Growth", Mkt\_Cap > 10000000000)  # Create scatter plot of market cap vs. growth for filtered data  ggplot(filtered\_data, aes(x = Coin, y = Mkt\_Cap, fill = Coin)) +  geom\_bar(stat = "identity", position = "dodge") +  scale\_y\_log10() +  labs(title = "Coins with Large Growth and High Market Cap", x = "Coin", y = "Market Cap ($)") +  scale\_fill\_manual(values = rainbow(length(unique(filtered\_data$Coin)))) +  theme\_bw() |

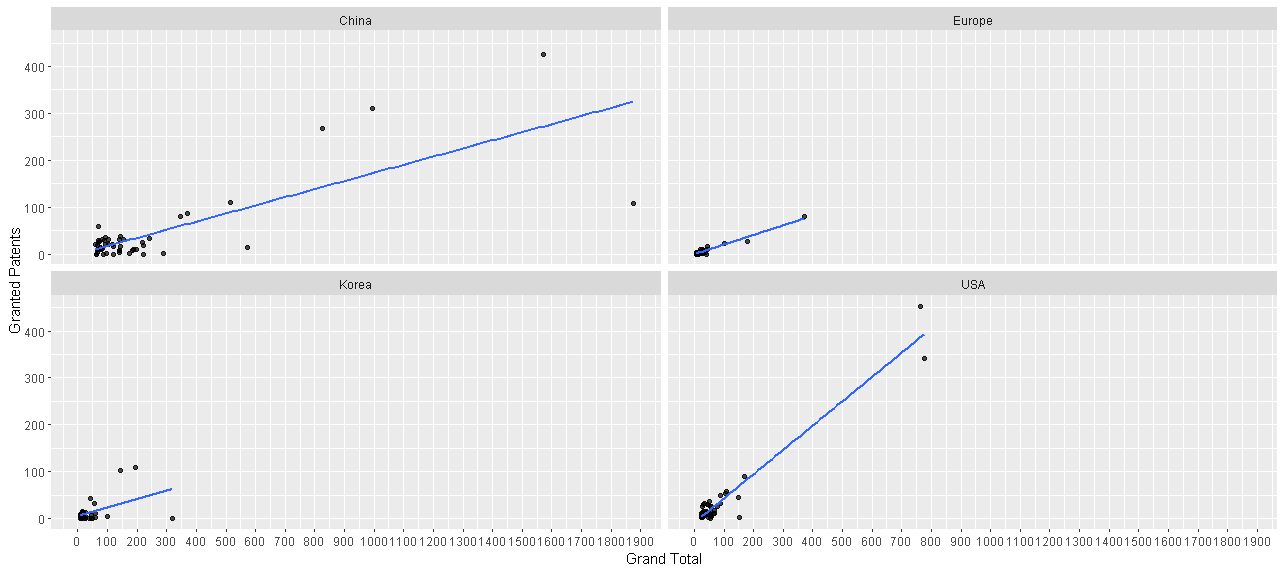
Shape, square

Description automatically generated

This Graph Shows the Top coin in crypto Market With Excellent Growth.

4)

|  |
| --- |
| ggplot(data, aes(x = Grand.Total, y = Granted.Patents)) +  geom\_point(alpha = 0.7) +  geom\_smooth(method = lm, se = FALSE) +  scale\_x\_continuous(breaks = seq(0, 2000, 100)) +  scale\_y\_continuous(breaks = seq(0, 1500, 100)) +  facet\_wrap(~ Country\_Origin) +  labs(x = "Grand Total", y = "Pending Patents") |



This Graph Shows the Country Who have a more focusing in Crypto Market. And China is the Most Granted Patent Holder.

5)

|  |
| --- |
| ggplot(data, aes(x = Grand.Total, y = Pending.Patents)) +  geom\_point(alpha = 0.7) +  geom\_smooth(method = lm, se = FALSE) +  scale\_x\_continuous(breaks = seq(0, 2000, 100)) +  scale\_y\_continuous(breaks = seq(0, 1500, 100)) +  facet\_wrap(~ Country\_Origin) +  labs(x = "Grand Total", y = "Pending Patents") |

A picture containing text, different, various, several

Description automatically generated

This Graph Shows the Country Who have a more focusing in Crypto Market. And China is the Most Pending Patent Holder in the World No Other Country is Even Close to Them.

6)

|  |
| --- |
| df\_sum <- data %>%  group\_by(Country\_Origin) %>%  summarize(Total\_Patents = sum(Granted.Patents)) %>%  arrange(desc(Total\_Patents))  # Get the country with the maximum granted patents  max\_country <- df\_sum$Country\_Origin[1]  # Create a pie chart with ggplot2  ggplot(df\_sum, aes(x = "", y = Total\_Patents, fill = Country\_Origin)) +  geom\_bar(stat = "identity", width = 1, color = "white") +  coord\_polar(theta = "y") +  ggtitle(paste0("Distribution of Granted Patents by Country (", max\_country, " has the most patents)")) +  scale\_fill\_viridis\_d() +  theme\_void() |

Chart, pie chart

Description automatically generated  
This Graph Shows That China is Holding Almost Worlds 50% Granted Patents in the World.

7)

|  |
| --- |
| data\_top <- data %>%  arrange(desc(Granted.Patents)) %>%  head(10)  ggplot(data\_top, aes(x = Company, y = Granted.Patents, fill = Country\_Origin)) +  geom\_bar(stat = "identity") +  ggtitle("Top 10 Companies by Granted Patents and Country Origin") +  xlab("Company") +  ylab("Granted Patents") +  theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  guides(fill = guide\_legend(title = "Country Origin")) |

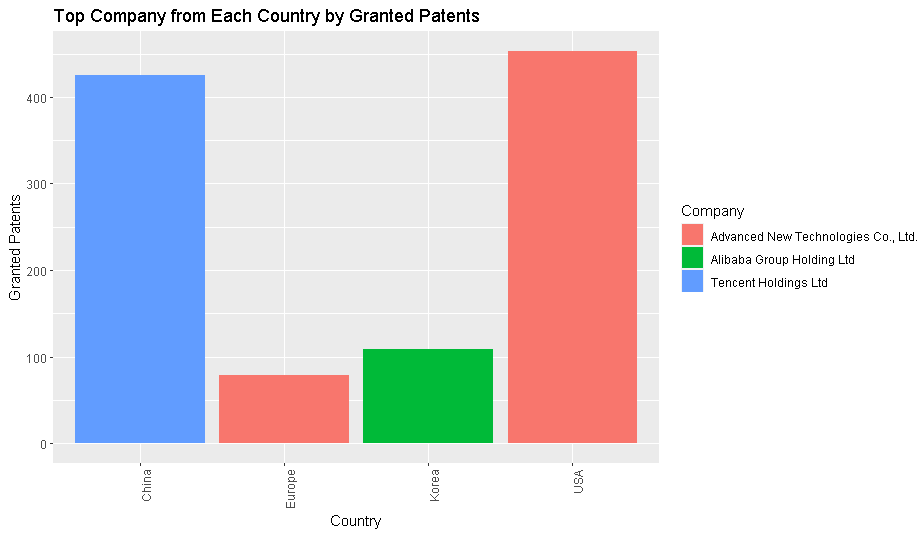
Chart

Description automatically generated

This Graph Shows That

8)

|  |
| --- |
| # Group data by Country\_Origin and get the top company in each group based on Granted.Patents  top\_company\_by\_country <- data %>%  group\_by(Country\_Origin) %>%  slice\_max(Granted.Patents) %>%  ungroup()  # Create a bar chart with ggplot2  ggplot(top\_company\_by\_country, aes(x = Country\_Origin, y = Granted.Patents, fill = Company)) +  geom\_bar(stat = "identity") +  ggtitle("Top Company from Each Country by Granted Patents") +  xlab("Country") +  ylab("Granted Patents") +  theme(axis.text.x = element\_text(angle = 90, hjust = 1)) |



9)

Code:

|  |
| --- |
| # Sort data by Total\_Patents  data\_by\_country <- data\_by\_country[order(data\_by\_country$Total\_Patents, decreasing = TRUE), ]  # Create a bar chart  ggplot(data\_by\_country, aes(x = Country, y = Total\_Patents, fill = Country)) +  geom\_bar(stat = "identity") +  ggtitle("Grand Total Number of Patents by Country") +  xlab("Country") +  ylab("Grand Total Number of Patents") +  theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) |

Chart, bar chart

Description automatically generated

**Shiny Dashboard Implementation:**

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, pie chart

Description automatically generated

Chart, pie chart

Description automatically generated

**Discussion and Conclusion:**

In This Analysis We can see that China will lead the future Blockchain Market. As they have the most granted patent, they have the most pending patent and also they have the most number of total patents. Most of the Top Block Chain Companies are from China. And they almost holds more than 40% of the total Patents of the World. And in the crypto market we can say that investor will invest more in the crypto market in the next day in 2 cryptos they are BNB and Cartano. As they have positive market growth and also they have a good market cap. So IT is a good sign. And market position is also good. And in the last 1h the market price variance is also good it has a low variance. The smaller variance tells that market position is better.