Project Report

# GitHub URL

**GitHub URL:**<https://github.com/oconnorbrian/UCDPA_BrianOConnor>

**Dataset:** <https://www.kaggle.com/datasets/shivamb/company-acquisitions-7-top-companies>

**Word Count: 2155** (*Excluding Figure Captions & References*)

# Abstract

In this report I discuss how I downloaded a dataset on Mergers and Acquisitions of big tech companies to gain a deep understanding of how these companies are spending billions of dollars on these acquisitions. I describe my process of importing, cleaning, and analyzing this data, illustrating my results, and gaining insights into when acquisitions are taking place, the different businesses being acquired, and the amount of money being spent.

# Introduction

I chose to analyze Mergers and Acquisitions of big tech companies as I am always seeing in the news about big tech companies acquiring large businesses with articles such as

“How Big Tech got so big: Hundreds of acquisitions” and “Exclusive: Big Tech Spent Over $30 Billion Acquiring Companies While Regulators Tried to Reel Them In”(Alcantara et al., 2021; DeGeurin, 2022).I decided it would be interesting to analyze data on previous acquisitions by these companies and gain insights into specific details of these acquisitions.

# Dataset

The dataset I chose for this project was “Merger and Acquisitions by Tech Companies” from Kaggle. This dataset provided a list of acquisitions made by the companies; Microsoft, Google, IBM, Hp, Apple, Amazon, Facebook, Twitter, eBay, Adobe, Citrix, Redhat, Blackberry, Disney. The dataset included the year and month of each acquisition, the name of the company acquired, the name of the parent company doing the acquisition, the cost of the acquisition and the business-use case of the acquisition, as seen in Figure 1. The source of this dataset was Wikipedia, CrunchBase and TechCrunch (Bansal, 2021).

Figure : First 5 Rows of Dataset

A screenshot of a computer screen

Description automatically generated with medium confidence

The dataset also included data on the country in which the acquisition was made, as well as the category and derived products of the acquisition. However, I chose to not use this data in my analysis as there was a large amount of missing data in each column, as see in Figure 2.

Figure : For Loop - Percentages of missing Data

Text

Description automatically generated

This dataset provided the necessary data for me to generate useful insights into mergers and acquisitions of some of the largest tech companies. With a total of 10 columns and 1455 rows, this dataset was within scope of the size required for this project. This dataset included both categorical and numerical data which would allow me to showcase the skills I have learned in this course.

I chose this data set as it was available for free on Kaggle with a Usability score of 10, which is the highest available score referring to Completeness, Credibility and Compatibility. With a relatively high 161 upvotes, this dataset was also proven useful by the userbase of Kaggle. This dataset was also updated annually, and as recently as 2021 which would allow for relevant and accurate analysis.

# Implementation Process

This section provides an overview of the entire implementation process of my project. Firstly, I downloaded my chosen dataset and imported the necessary python packages, as seen in Figure 3, it was then necessary to import the dataset into my notebook, JupyterLab.

Figure : Importing Libraries

Text

Description automatically generated

I transferred my dataset into the same folder as my notebook which allowed me to easily import the csv file using pandas, without having to specify the file path, as seen in Figure 4. Using pandas, I was able to also create my DataFrame easily in the same line of code. I then checked the data was imported correctly, viewing the first and last 5 lines of code and examining the DataFrames information which showed the correct 10 Columns and 1455 Rows.

Figure : : Importing Data from flat file - CSV

Graphical user interface, text

Description automatically generated

Next, I retrieved data using an API. I opted for this alternate method of importing as I did not have access to relevant online SQL, and web scraping is less relevant for business use-cases. I used the free ClearBit logo API because it would allow me to get logos for all companies in my dataset, that although out of scope for this current project, I could use in future as part of my visualizations (ClearBit, 2023). I used the requests.get() method to retrieve the API and used conditional statements to verify the requests success, as seen in Figure 5.

Figure : Retrieving Data using API

Text

Description automatically generated

Having already created my pandas DataFrame, as seen in Figure 4. I then cleaned my data by sorting, Indexing, and grouping my data, checking for duplicates, and replacing missing values. I decided to first sort my data by putting Parent Company in alphabetical order, and by their most recent acquisitions allowing for ease of viewing the latest acquisitions of each company, as seen in Figure 6.

Figure : Sorting DataFrame

A screenshot of a computer

Description automatically generated with medium confidence

Next, I checked for duplicates in the Acquired Company and ID columns as these were the only columns required to be unique. I did this in two ways, firstly counting duplicates which presented that there were no duplicates (False) in all 1455 rows of the DataFrame, and secondly by creating a test DataFrame, dropping duplicates and seeing that it did not impact the number of rows, As seen in Figure 7.

Figure : Checking for Duplicates

Text

Description automatically generated

Having established that the ID column was unique, I then set ID as the index of the DataFrame, as seen in Figure 8.

Figure : Setting Index

A screenshot of a computer

Description automatically generated with medium confidence

Next, I replaced the missing values “-“ and “undisclosed” with 0. I chose this method as it was simple and allowed me to have numeric values for Acquisition Price and Acquisition Year. I then chose to have 0 for all missing values for consistency. Using the replace method instead of the fillna method allowed me to replace “undisclosed” with ease as it is not NaN. I then printed a description of my cleaned dataset, as seen in Figure 9.

Figure : Replacing Missing Values

A screenshot of a computer

Description automatically generated with medium confidence

Finally, I created a new DataFrame “df2” to display an example of grouping by business, as seen in Figure 10, as I was happy with how my original DataFrame “df” was already sorted.

Figure : Grouping Example

A screenshot of a computer

Description automatically generated with medium confidence

I began my analysis by using a for loop to display the percentage of rows that have 0 information in each column, as seen in Figure 2. I then converted the numeric values from objects to int64, as seen in Figure 11.

Figure : Object to int64

Text

Description automatically generated

I used NumPy functions to gain insights into the amount of money (USD) being spent on acquisitions, calculating the total money spent on all recorded acquisitions, the highest and lowest recorded amounts on single acquisitions and the average amount spent on an acquisition, as seen in Figure 12. I created a list of the prices for use in calculations, opting for a list over a dictionary as lists allow for duplicates and associated keys were not necessary. I also used conditional statements to get the lowest amount and created a custom function to get the average.

Figure : NumPy Functions

Text

Description automatically generated

Next, I analyzed which companies have had the highest number of acquisitions by creating a DataFrame of the number of Parent companies and renaming the columns. I then plotted this data on a barplot, as seen in Figure 13. I opted for a larger and bold font for the title to make it stand out, as well as ensuring I labelled both axes. I removed the trim as I believe it looked cleanest without and chose the colorblind palette because I am colorblind, and it helps to distinguish the colors between each bar. I kept this style of graph consistent for all my graphs, adjusting the style only for size depending on the graph.

Figure : Creating new DataFrame.

Text

Description automatically generated

I then analyzed the most popular businesses being acquired by both Apple and Microsoft by defining a custom function that locates an inputted company and then repeating the previous steps while using this function to create my initial DataFrame, as seen in Figure 14.

Figure : Custom Function

Text

Description automatically generated

I then Merged these two DataFrames, opting for an outer join to collect every record, and setting both “left\_on” and “right\_on” to the desired columns, although as these columns had the same names it would have been possible to just specify “on”, as seen in Figure 15.

Figure : Merging DataFrames

A screenshot of a computer

Description automatically generated with medium confidence

I also Concatced the top 5 rows of these DataFrames together, setting the index and sorting by largest totals to compare the most common acquisitions of both companies. I then created a bar chart to illustrate this comparison, as seen in Figure 16.

Figure : Concat DataFrames

A screenshot of a computer

Description automatically generated with medium confidence

The next step of my analysis involved finding and plotting the average amount each company spent on an acquisition. I grouped by Parent Company and got the average price. I displayed this as a table and then plotted it as a bar chart, as seen in Figure 17.

Figure : Groupby Column

Text

Description automatically generated

However, as Disney had a much higher average than the other companies, I removed it by creating a maximum average of less than 3,000,000,000 USD, and plotted the remaining companies as seen in Figure 18.

Figure : Remove Disney with ">"

Text

Description automatically generated

Next, I analyzed the relationship between acquisitions and months/fiscal quarters. I organized the months in order of their occurrence in the year rather than alphabetically and plotted them in a histogram as seen in Figure 19.

Figure : Ordering Months

A screenshot of a computer

Description automatically generated with medium confidence

I also created a map to further organize each month into their respective fiscal quarters and plotted the acquisitions each quarter on a bar plot, as seen in Figure 20.

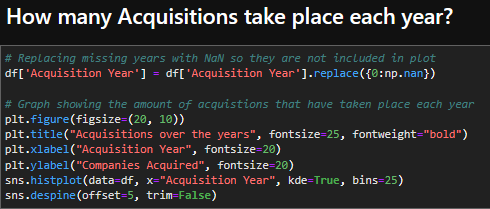
Figure : Ordering by Quarter

Text

Description automatically generated

Finally, I analyzed how many acquisitions took place each year, I replaced all 0 values with NaN as seaborn would only include the numeric data in the histogram, as seen in Figure 21.

Figure : Replacing 0 Values



I then chose to further analyze a specific business-use case, replacing all Artificial Intelligence related business use-cases with a generic “AI” value, and plotting the number of AI business acquisitions per year, as seen in Figure 22.

Figure : Replacing AI values.

A screenshot of a computer

Description automatically generated with medium confidence

# Results

In this section, I will illustrate the results of my analysis and describe what information they provide. Firstly, my analysis on acquisition prices provided the following results illustrated below in Figure 23. I created strings to describe what each number result was being referred to, including the dollar sign to show the unit.

Figure : Money spent by companies on acquisitions.

Text

Description automatically generated

I then looked at which companies had the highest number of acquisitions. Here the bar chart outlines each parent company and their total acquisitions. We can see Microsoft and Google have the highest total acquisitions, whereas Blackberry and Disney have the lowest, as seen in Figure 24.

Figure : Company Total Acquisitions.

Chart, bar chart

Description automatically generated

Next, I looked at the most popular type of businesses being acquired by both Apple and Microsoft, here we can see that Software is the most popular business acquired by both Apple and Microsoft, however their other most popular businesses to acquire vary greatly, as seen in Figure 25.

Figure : The most popular type of Businesses being acquired by both Apple and Microsoft?

Chart, bar chart, funnel chart

Description automatically generated

I then combined the top 5 most popular business to acquire of each company as seen in Figure 26. I did struggle to make a visual distinction between the companies on the graph, however I have included my attempts at such in my code.

Figure : Microsoft and Apples Most Popular Business to Acquire.

Chart, histogram

Description automatically generated

Next, I looked at the average amount each company is spending on an acquisition, where we can see in Figure 27 Disney spends much more on average than other companies. I also included Figure 28 which excludes Disney and gives a better comparison of the remaining companies, allowing us to see things such as IBM and Microsoft spending a similar average more clearly.

Figure : Average Amount of Money Spent on an Acquisition (including Disney)

Chart, bar chart

Description automatically generated

Figure : Average Amount of Money Spent on an Acquisition (excluding Disney)

Chart, bar chart

Description automatically generated

I illustrated the acquisition per month in Figure 29, displaying November as least acquisitions and June as most. I also displayed the acquisitions by quarter, using the neutral color blue and positive green to showcase the higher acquisitions in Q2 and Q3, as seen in Figure 30.

Figure : Acquisitions Per Month

Chart, bar chart

Description automatically generated

Figure : Acquisitions Per Fiscal Quarter

Chart, bar chart

Description automatically generated

Finally, I illustrated the total acquisitions over the years, which displayed a large uptrend in the late 2000s and early 2010s as seen in Figure 31. Similarly, I illustrated the total Artificial Intelligence business acquisitions in Figure 32, which shows they only begun being acquired in the mid-2010s. I believe there is potential for machine learning to help predict the number of future Artificial Intelligence business acquisitions that will happen in a year for these large tech companies using regression methods, this could give a company insights into the level of competition for acquisitions in future years.

Figure : Acquisitions over the Years

Chart, histogram

Description automatically generated

Figure : Artificial Intelligence Company Acquisitions

Chart, histogram

Description automatically generated

# Insights

* There is a large range of money being spent on acquisitions, from as little as $200,000 for an acquisition up as far as $7,130,000,000 and with the recorded prices of these companies acquisitions alone accounting for over half a trillion dollars ($519,877,989,500) in expenditure, as seen in Figure 23.
* Microsoft has had the highest number of acquisitions out of all companies listed. With Microsoft’s most popular acquisitions being software businesses, while also acquiring a large amount of video game businesses.
* Disney has the lowest number of acquisitions out of all companies listed, however they spend 7.5 times more money on an average acquisition than the nearest company listed.
* For technology companies, there has been a larger number of acquisitions in Q2 and Q3, than Q1 and Q4, with June being the month where most acquisitions take place. However, not all acquisitions under $92 million are reported, so the actual distribution may be different (Vynck & Zakrzewski, 2021).
* Total Acquisitions of these companies was steadily increasing from the 1990s to the early-2010s, with a plateau in the mid to late 2010s. However, there was a further steep drop off in 2020, which I believe could be related to the beginning of the Covid-19 Pandemic and companies saving their cash reserves (Zhou et al, 2022).

# References

Alcantara, C. *et al.* (2021) *How big tech got so big: Hundreds of acquisitions*, *The Washington Post*. WP Company. Available at: https://www.washingtonpost.com/technology/interactive/2021/amazon-apple-facebook-google-acquisitions/ (Accessed: January 13, 2023).

Bansal, S. (2021) *Merger and acquisitions by Tech Companies*, *Kaggle*. Available at: https://www.kaggle.com/datasets/shivamb/company-acquisitions-7-top-companies (Accessed: January 15, 2023).

*Clearbit* (2023) *Clearbit API Documentation For Developers*. Available at: https://dashboard.clearbit.com/docs#logo-api (Accessed: January 30, 2023).

DeGeurin, M. (2022) *Exclusive: Big tech spent over $30 billion acquiring companies while regulators tried to reel them in*, *Gizmodo*. Gizmodo. Available at: https://gizmodo.com/amazon-google-apple-antitrust-meta-acquisition-1849546460 (Accessed: January 30, 2023).

Vynck, G.D. and Zakrzewski, C. (2021) Tech Giants quietly buy up dozens of companies a year. regulators are finally noticing., The Washington Post. WP Company. Available at: https://www.washingtonpost.com/technology/2021/09/20/secret-tech-acquisitions-ftc/ (Accessed: January 13, 2023).

Zhou, D., Zhou, H., Bai, M. and Qin, Y., 2022. The COVID-19 outbreak and corporate cash-holding levels: Evidence from China. *Frontiers in Psychology*, *13*.