

Visualising Electric Vehicle Demographics in the USA

Charlie George Barrow

1 PROBLEM STATEMENT

Sustainable investing has generated an enormous amount of interest over the past two years. With USD 38 trillion assets under management (AUM) within portfolios in 2019 (Yahoo, 2021) ^[1]. Additionally, sustainable investing in Europe alone is projected to reach €7.6tn over the next five years, outnumbering conventional funds, due to investors' focusing more on sustainable investing (Financial Times, 2020) ^[2]. A key sector within sustainable investing, is the Electric Vehicles (EV) sector, where companies such as Tesla, has over the last two years has witnessed its stock price increase over 1,700% (Bloomberg, 2020) ^[3]. Hence, the foundations of this report aims to visualise the EV sector within the USA and answer the following analytical questions;

- 1) How the EV demographic has changed over time.
- 2) What brands/models are denominating this sector.
- 3) How the EV battery technology of has evolved and what models/brands are ahead of the technology curve.

Answering these questions should potentially suggest the prospective movements of this sector and not only for notable EV companies such as Tesla, General Motors and Fords stock prices. Also, other sectors whose growth is linked to EV's including; Graphite, Cobalt and lithium-ion production. Overall, through manipulating an US EV Population Dataset ^[4] containing each registered EV from years 1993-2021, will help solve the analytical questions. Whereby, exploiting the datasets attributes; EV's geographical location that can be used to assess the demographics. Brand and Model Type to assess supremacy. Lastly, miles per charge (MPC) to analyse EV battery capacity evolution over the last 28 years.

2 STATE OF THE ART

Podorozhnyy, 2020 ^[5] published a paper that developed new data visualizations in the electric mobility sector in the US deriving from the data obtained by McKinsey, 2020 ^[6] that highlighted consumer EV awareness and purchase actualisation, alongside EV OEM (Original Equipment Manufacturer) pipeline for years 2019-2025. Additionally, researching what advancements in battery technology had taken place in the EV sector, alongside how successful has US EV deployment and adoption been and what is the status of EV mass-market adoption. Furthermore, such techniques and approaches used in this comparative study focused on data manipulation by applying Ideation techniques such as Brainstorm, Brainwrite, and Worst Possible Idea. Resulting in, manipulating the data into market analysis visualised

through bubble map charts and histograms. This not only relates to this report which similarly planned to use Python's Matplotlib.pyplot library and create a histogram to visualise the ranking of EV Models based on their battery capacity to understand the EV battery evolution. Alongside, improving upon the study through combining, Python's Plotly/Geopandas library to create a geographic bubble chart visualising current EV mobility in the USA. Similarly, to Podorozhnyy, 2020 ^[5] when visualizing EV demographics and advancements in technology. Each papers result could be compared later on how they correlate. Such as, comparing Podorozhnyy, 2020 ^[5] discovery that the California state and city San Francisco dominated the US for most EV in a city with 19.5% of all EV registrations.

Secondly, a paper published by H.A. Ayad et. al, 2020 ^[7] developed an Interactive data visualization platform to present the adaption of EV's in Washington, California and New York. Who, through using the tool Microsoft Power BI, assessed the number of EV's and specified their types, brand, and models used in each state. Furthermore, the approach involved using Microsoft Power BI which is an analytics service provided by Microsoft, aiming to provide users with interactive visualizations and business intelligence capabilities, alongside an interface simple enough for end-users to create their reports and dashboards. So, whilst the report did not use the approach of the Python language to produce the visualisations, its results however once more offer a comparative aspect that can be compared with this study. For example, H.A. Ayad et. al, 2020 ^[7] discovered the difference in number of EVs in each state, stating that California is the highest state that uses EVs, followed by New York and Washington. Plus, discovering Tesla to be the top producer of Battery Electric Vehicles (BEV). These findings can be used to compare with this reports results, as it coincides with its research questions. Moreover, H.A. Ayad et. al, 2020 ^[7] used the following visualisation tools to obtain these results; bar charts, line charts and geographic maps. Yet again highlighting a comparative aspect, as this study likewise creates the same visualisations across the US using a geo-map, as well as, using histograms, bar and line charts to visualise similar research questions, Including, which brands are denominating this sector and which state produces the most EV's.

3 PROPERTIES OF THE DATA

The EV raw data set ^[4] contained 15 attributes with 51493 instances detailing all EV registrations and their specifications in the USA from the years 1993-2021. The relevant attributes fall to 11 that includes County, City State, ZIP Code, Model

Year, Make, Model, Electric Vehicle Type, Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Range (Miles) and Coordinates. County, City State, ZIP Code and their registered coordinates. The Model and Year columns looks at each EV brand, different models and the year they were released. Total models in this database tally to 51. Similarly, the make column highlights the different brands registered corresponding to the EV models. Total models in this database tally to 24. Additionally, there is the Electric Vehicle Type column that details the car battery specification, conveying how much the battery relies on either pure electricity or fossil fuels, since some EV's are hybrid and have a combination of both. Alongside this, there is the Clean Alternative Fuel Vehicle (CAFV) eligibility, which essentially classes a vehicle by law if it is considered a fully Clean Alternative Fuel Vehicle or if it isn't. This is a crucial column, since those who do not meet the criteria will be filtered in order to improve the datasets quality therefore, signaling the importance of data pre-processing before the data itself can be visualised. Lastly, there is the Electric Range (Miles) attribute, which as the name suggests, is the electrical range of the vehicle per charge/full-tank that is measured by miles. This will be used to understand each model's battery range, alongside how battery capacity has changed historically.

Overall, the model year and Electric Range (Miles) attributes specified for each EV, makes temporal analysis achievable, giving the full scope of database and answers the research questions; how EV model's technology has evolved over the databases periods. Together with, how the number of registered EV's have changed over time which is visualised in *Figure 1* showing exponential growth in registered vehicles in the database during the last two decades. Additionally, temporal analysis can also be displayed through what brand and its respective models dominate each city and state across the USA at a given period.

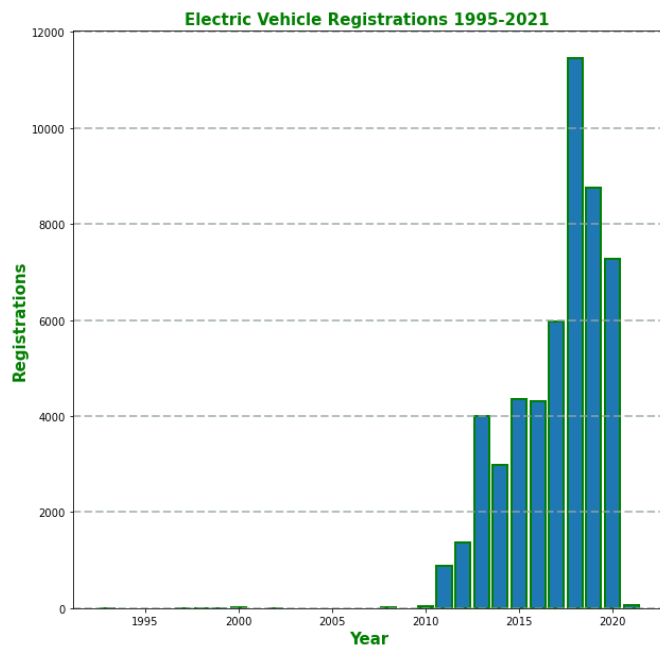


Figure 1: USA Electric Vehicle Registrations.

In addition, spatial analysis is also possible through the longitude/latitude coordinates attribute column provided for each registered EV, making a visualisation on a geographic map achievable as displayed in *Figure 2* that underlined the State of Washington to have the most registered EV's in the country. Also, the Zip codes are correspondingly provided; however, it will be imperative to make sure these pair up with the coordinate's variables, making sure each column does not contain any missing variables or highlights a different area. In summary, for analysis, this study centers around how the EV demographic has changed over time in the US, what brands/models are dominating the US, and how the battery technology of EV's has evolved since the databases inception.



Figure 2: USA Electric Vehicles Demographics.

4 ANALYSIS

4.1 Approach

Diagram 1 represents the decisive approach to answering the analytical questions, including the analytical stages, their purpose and methods applied. Below further examines details of each stage, including the relationship between human reasoning and computational methods applied. Together they are crucial for each analytical stage and only an effective combination of both will generate comprehensive analysis and results.

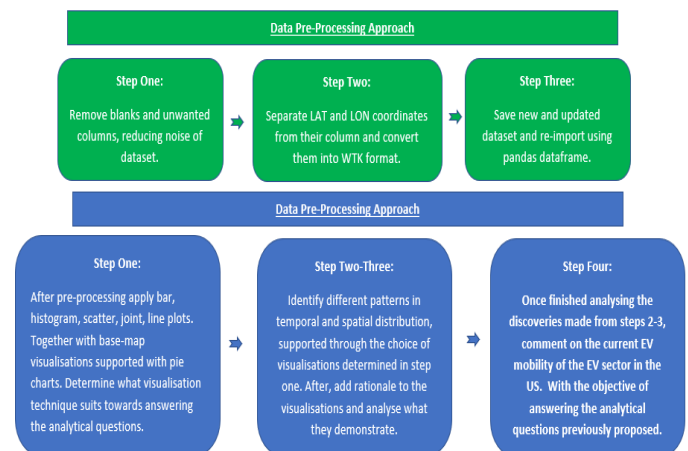


Diagram 1: Approach to visualisation

Data Processing:

Data processing commenced with the removal of immaterial attributes that held no value in the database, including the 'Legislative District', 'DOL Vehicle ID', 'Base MSRP' and 'VIN (1-10)' attributes, which was done in order to level down the noisiness of the dataset. Afterwards, came the removal of the non-uniformity EV model spec data that didn't meet the Alternative Vehicle Fuel Spec that legally classes an automobile as an EV. Achieved by filtering the string 'not eligible due to low battery range' from the column 'Clean Alternative Fuel Vehicle (CAFV) Eligibility'. Additionally, the "Electric Range (Miles)" was sorted by its values from highest to lowest, in order to enhance the visual quality and understanding of what the attribute was displaying. Highlighting that Tesla's Model S had the most long-lasting EV battery. Also, the dataset itself contained one erroneous tag due to a missing variable in its coordinate's column. At first this went undetected, however when applying the coordinates in a WKT format so that the Python library Geopandas could interpret, it delivered a value error. Nevertheless, after removing the missing value, it allowed the database to run efficiently. Although, a further issue developed when discovering the WTK format was limiting the application of the converted coordinates to other geographic map-based Python libraries, such as Plotly and Pydeck. To combat this, the coordinate's data column was split into lateral and longitude columns so they could be applied to these alternative libraries. Plus, were interchanged around since they were miss placed in the original dataset showing lateral where longitude should have been.

STAGE ONE:

Use Python's panda's library conduct initial data exploration on the pre-processed data applying bar, histogram, scatter, joint, line plots. Together with base-map visualisations supported with pie charts. This stage is anticipated to result in determining what proficient visualisation techniques are relevant in answering the analytical question.

STAGE TWO- THREE:

Identify patterns using temporal and spatial analysis. Supported by bar, histogram, scatter, joint, line plots created. Together with, base-map visualisations, supported by pie charts produced in stage one. Afterwards, through research and general knowledge, explain how the visualisations relate to social economic behavior, including US climate policies on global warming and greenhouse gases. In doing this, human reasoning will help to bring the results back into context.

STAGE FOUR:

Through analysing the discoveries from stages 2-3, this stage intends to comment on the current mobility of the EV sector, through summarising EV brand and model ownership in the US, what state houses the most EV's, ranking EV based on their battery performance, and the correlation between the years and EV battery capacity growth.

4.2 Analysis Process

Temporal Analysis

EV Growth 1993-2020

From analysing *Figure 1*, that was constructed using Python's matplotlib library combined the databases Model year and count attributes. These variables were selected since they would specify how the EV demographic had changed over time, reflected through the change in EV registrations year on year. As such, *figure 1* displayed a clear upward time-trend, exhibiting a surge in in EV registrations from the years 1993-2021. Immediately signaling that e-mobility is becoming more adopted in the US and moving away from traditional fossil fuel automobiles.

Furthermore, the year 2010 onwards in particular, demonstrated exponential growth, highlighting an EV register increase of 139070.27%. The most EV registrations occurred during the last three years starting at 2018 with 11454 followed by 2019 with 8766 and 2020 with 7270. Such results do not come as a surprise, since in late 2009, congress passed the "The American Recovery and Reinvestment Act of 2009", that established tax credits for consumer's that purchased electric vehicles (HEV TCP, 2009) ^[8]. This was the first material law passed in the US catering towards tackling carbon emissions produced by vehicles, since it was beneficial for national security (David et al, 2009) ^[9]. Since then, numerous amendments have occurred to this law, alongside new ones passing to encourage the adoption of the EV's. Additionally, the peak years and slow decline of EV registrations from 2018-2020 can be rationalised during the Trump Presidency, whom in the midst of his term, took a cold stance against renewable energy products. Demonstrated from when he reversed the Obama-era efficiency standards, vacated from the Paris agreement, placed higher import tariffs on renewable energy products and cut the US's funding for tackling climate change (SEIA, 2018) ^[10]. Subsequently, such actions drove up EV production prices dramatically where originally, they were anticipated to fall, which for the time put off consumers purchasing EV's (Energycentral, 2020) ^[11].

EV Battery Technology 1993-2020

From analysing *Figure 3*, which was constructed using the Python's visualisation library Seaborn and its jointplot script from its website ^[12] combined the databases attributes Model year and Miles Per Charge (MPC). Such variables were selected since they would effectively display the correlation between the databases years and the transformation of EV battery capacity during the period. Yet again, this chart displayed a clear upward time-trend, exhibiting an intense evolution in EV battery capacity in the MPC variable from the years 1993-2021.

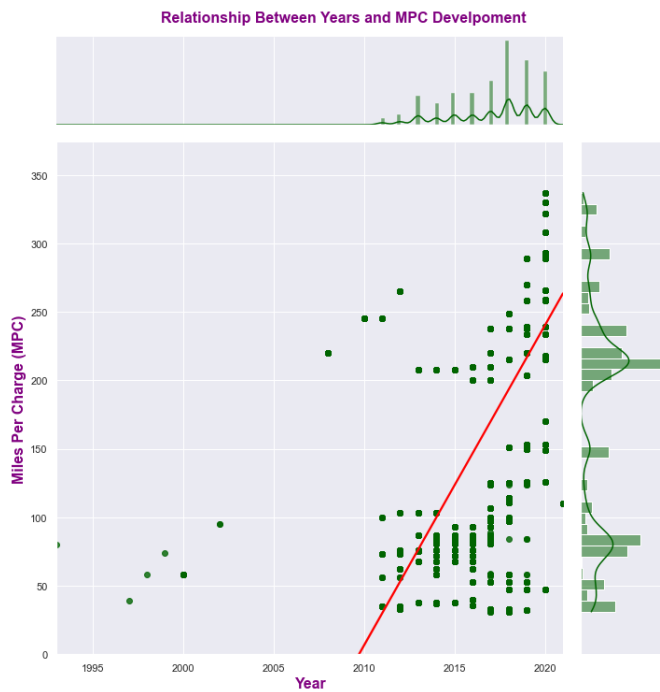


Figure 3: USA Electric Vehicle Battery MPC Evolution

Moreover, the chart exhibits a clear positive relationship with the progression of time and advancements in EV battery capacity, as witnessed from 1993. The max battery capacity at the time stood at 31 MPC. Since then, EV battery has increased 987.10% and now stands at 337 MPC as of 2020. To put this into perspective, in 1993 on a single charge, an EV could take you up and down the length of Manhattan once. Demonstrating EV at the time to be a limited city vehicle. However, presently in 2020 the most advanced EV can travel from New York to Washington and still have over 100 miles left before requiring charge. Overall, demonstrating significant improvements to EV technology. Plus, through analysing the chart's regression line MPC, it is likely going to increase further in the future. This result can be rationalised through how the EV battery technology has changed during this period. Which according to Ahmed et al (2016)^[13] developments and adaption of Lithium-ion batteries by OEM's has been revolutionary in creating far more efficient EV's. As Lithium-ion batteries contain far greater efficiency for storing and discharging electricity, compared to other materials. Alongside, being rechargeable and sustainable, they can last for well over 10 years before needing to be replaced^[13].

Spatial Analysis

Distribution of EV's Across the USA

From analysing Figure 2 and 4, 2 was produced using Python's Plotly library and its website script^[14] to plot all EV coordinates on a geographic map using a public USGS imagery map. Similarly, 4 was produced using Python's visualisation library Pydeck script^[15] to plot all EV coordinates in further detail to visualise EV distribution by state/city using its HexagonLayer and 3D imagery. Therefore, using the coordinates variables allowed to visualise how EV

demographic has changed over time by each state. And in summary, Figure 2 and 4 demonstrates how EV registrations in the database are highly condensed in the North Western United States and very little registrations have occurred outside of this area.

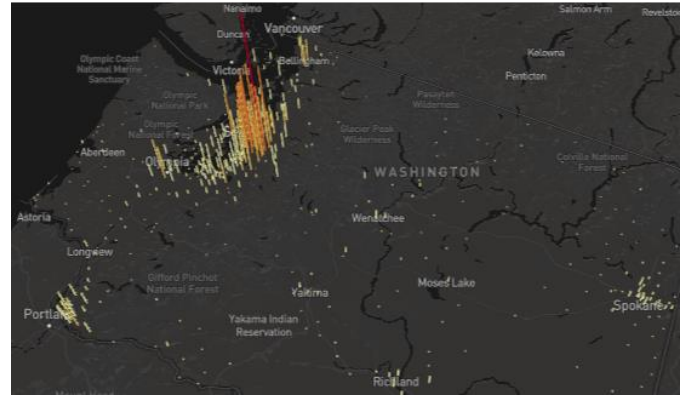


Figure 4: USA Washington Electric Vehicles Demographics

Additionally, Figure 2 depicts the state of Washington containing the most EV registrations across the US, this is thought-provoking since this finding contradicts H. A. Ayad et. al, 2020^[7] who established instead the state of California surpassed Washington as most EV registrations. The rational of these finding isn't truly clear; it could only be narrowed down to that one of the datasets may contain insufficient data and doesn't truly reflect the exact EV registrations in the US at the time. Finally, Figure 4 portrays within the Washington State, that the City of Seattle contained the most registrations of EV out of all US cities, followed by Bellevue and Redmond. These findings can be explained through the development of the comprehensive EV policy packages passed in Washington, whom have been employing their own EV promotion policies^[16]. Since 2010 Washington has been reducing EV consumer purchase barriers through policies, incentives, infrastructure buildout, and awareness campaigns for EV's (ICCT, 2019)^[16]. Together with, adopting California's zero-emission vehicles (ZEV) regulations to help catalyse the market, expand model availability, and provide assurance for charging infrastructure investments^[16].

EV Brand and Model distribution in the USA

Figures 5 and 6 were both produced using Python's visualisation library Seaborn and its barplot script from its website^[17] utilizing the Make and model variables in the database, to help visualise what brands and models are denominating the EV marketplace in the US. Moreover, the figures themselves underlined that the most prominent EV brands and models were Tesla, Nissan and Chevrolet, whom demonstrated the most presence in the US, through EV registrations.

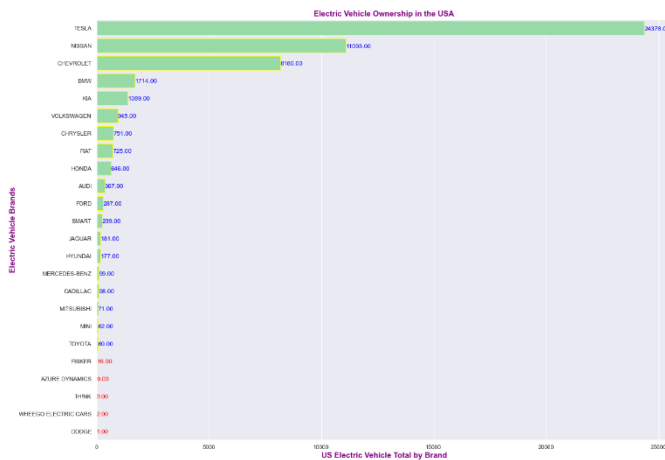
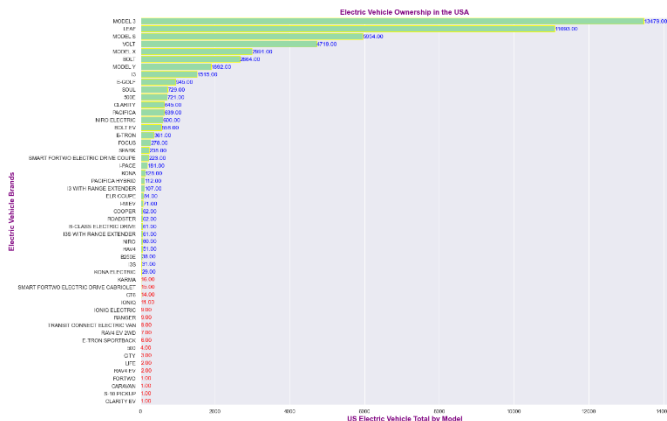


Figure 5: EV Brand Distribution in the USA

Figure 5 confirmed that Tesla as a brand totals virtually half of all EV registrations in the US with 24378, followed by Nissan trailing behind with 11093 then Chevrolet with 8180. Expectedly, in Figure 6 Tesla's Model 3, Nissans Leaf series, and Chevrolet volt series demonstrated the highest in model popularity in the US. What is noteworthy is Hyundai features 14th regarding EV brand ownership in the US, but yet for EV model choice in the US it is sixth. Largely, it is clear Tesla is head and shoulders above the rest which narrows down to firstly brand loyalty since it can be compared as to how Apple's dominated the smartphone industry through its brand loyalty over the years. Similarly, Tesla and Apple both have excellent designs and engineering qualities through using only the finest materials and hiring highly-skilled technologists to create their products which in turn gives them top performance. In particular for Tesla seen in Figure 7 it displays the highest battery pack that can travel the farthest out of its rivals. Overall, such brand loyalty has seen the carmaker in a consumer report obtain 98 out of 100 points concerning the driving experience, comfort, value, and styling alongside things such as audio and climate systems (Business Insider, 2019) [18].



Fourthly, finding that Tesla accounts for half of all EV registrations in the US (24378) with its Model 3 being favored. Followed by, Nissan's Leaf series (11093) then Chevrolet's Volt series (8180). Lastly, concluding that Tesla dominates the EV sector regarding its battery capacity with its Model 3 able to travel 337 miles on a single charge. Overall, its clear e-mobility is gaining momentum and is likely going to continue growing during the next decade. Since, with the Trump administration ending and Biden's administration commencing who states will reverse Trumps anti-global warming antics (The Guardian, 2021) [21]. Thus, allowing a smoother transition from fuel-based vehicle's to EV's, having a positive impact not only EV companies also Graphite, lithium-ion battery sectors.

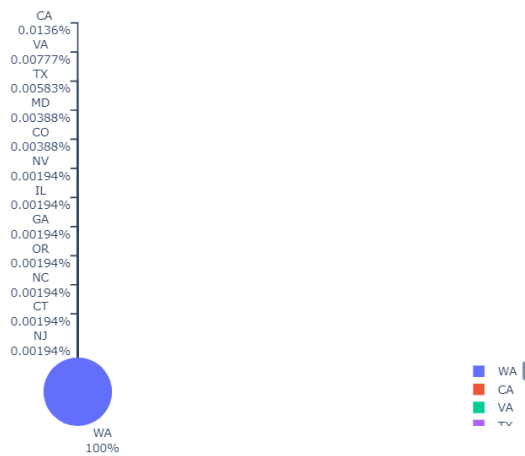


Figure 8: USA EV Registrations by State

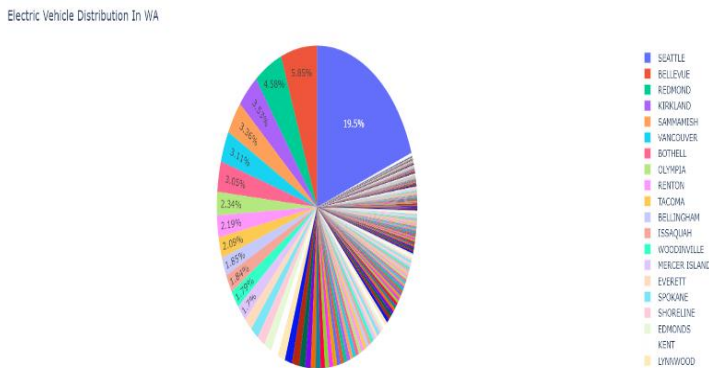


Figure 9: USA EV Registrations in Washington

5 CRITICAL REFLECTION

In summary, this project effectively pre-processed the dataset and created an interpretable one that made sense of the information that it was concealing. However, the dataset quality in the future could be improved. Since it was evident that when comparing this report's discoveries to that of H. A. Ayad et. al, 2020 [7] there were clear gaps in data from both

reports that didn't reflect true EV registrations in the US. This could have been tackled with combining these datasets and removing overlapping data, which could have given a richer insight into EV distribution across the US and created more needed noise in the dataset.

In addition, the report was effective in using both temporal and spatial analysis to answer its original problems statements it set out. In particular, using temporal analysis to identify that the EV sector in the US is gaining traction as highlighted through its explosive rise in registrations shown in *figure 1* from years 1993-2020. Additionally, through this paper's rationalisation of current political events in the US. It established that EV registrations can be anticipated to rise in the future as old and new EV tax incentives are restored and passed under the Biden Presidency. Secondly, this paper through using temporal analysis successfully identified a pattern in EV technological advancements revolving around battery capacity from 1993-2020. Alongside, rationalising how lowering of manufacture costs that will drive down prices. Nevertheless, the temporal analysis applied could have been enhanced. Upon reflection, whilst this paper successfully visualised the demographic change in EV registrations and distribution, it failed to highlight how each state and cities individual EV registrations have changed over time. Which could have been possible using a map subplot script from Python's Plotly library [22] that creates small map multiples in Python to help visualise changes in a geographic area over time.

Regarding spatial analysis, this report successfully identified which brands and models are dominating the EV sector in the US and by what margins. Stating that Tesla, Nissan and Chevrolet as the dominant brands with their respective models. However, this spatial analysis could have been improved through using temporal analysis to identify how EV brand and model ownership has changed over time. Together finding out on a by city or state basis which brands or models dominate the area, which could be of value for the automotive companies previously stated, regarding installation of charging stations and repair shops. Finally, its clear there is an opportunity present for climate investing for the future within EV's and its raw materials. Which could be a potential long-term opportunity for prospective investors who wish to seek a strong return in this emerging market, however still want to maintain or adopt a sustainable investing strategy in their portfolio.

Table of Word Counts:

		Word Count
1	Problem statement	250
2	State of the art	500
3	Properties of the data	500
4.1	Analysis Approach	500
4.2	Analysis Process	1555
4.3	Analysis Results	200
5	Critical reflection	459

REFERENCES

- [1] Yahoo Finance (2020), Changes in ESG investing [Online], Available at: <https://uk.finance.yahoo.com/news/global-environmental-social-governance-esg-102200058.html> (Accessed 1 January 2021).
- [2] Financial Times (2020) ESG Fund Projections [Online], Available at: <https://www.ft.com/content/5cd6e923-81e0-4557-8cff-a02fb5e01d42> (Accessed 1 January 2021).
- [3] Bloomberg (2020) Tesla Share Price Gains [Online], Available at: <https://www.bloomberg.com/news/articles/2020-12-31/from-tech-to-tesla-stocks-get-14-trillion-boost-in-covid-year> (Accessed 1 January 2021).
- [4] Kaggle US EV Population Dataset (2020) [Online], Available at: <https://www.kaggle.com/vijayakishoredusi/evpopulation> (Accessed 1 December 2020).
- [5] Podorozhnyy, Gleb (2020) Data visualization of the current state of electric mobility market [Online], Available at: <http://essay.utwente.nl/83174/> (Accessed 1 December 2020).
- [6] McKinsey & Company E-mobility Report (2020) [Online], Available at: <https://www.mckinsey.com/~/media/mckinsey/industries/automotive%20and%20assembly/our%20insights/the%20road%20ahead%20for%20e%20mobility/the-road-ahead-for-e-mobility.ashx> (Accessed 1 December 2020).
- [7] H. A. Ayad et. al, Developing an Interactive Data Visualization Platform to Present the Adaption of Electrical Vehicles in Washington, California and New York (2020) [Online], Available at: <https://ieeexplore.ieee.org/abstract/document/9101928> (Accessed 1 December 2020).
- [8] HEV TCP, EV Laws (2010) [Online], Available at: <http://www.ieahev.org/by-country/united-states-policy-and-legislation> (Accessed 1 January 2021).
- [9] David B. Sandalow, Plug-In Electric Vehicles: What Role for Washington? (Washington: Brookings Institution Press, 2009). (Accessed 1 January 2021).
- [10] Trumps War with Solar Energy (2018) [Online], Available at: <https://www.seia.org/initiatives/climate-change> (Accessed 1 January 2021).
- [11] Detrimental effects of tax increases for renewables for consumers (2020) [Online], Available at: <https://energycentral.com/news/end-tax-credits-solar-and-wind-power> (Accessed 1 January 2021).
- [12] Python's visualisation library Seaborn's jointplot script (2020) [Online], Available at: <http://seaborn.pydata.org/generated/seaborn.jointplot.html> (Accessed 1 December 2020).
- [13] Pesaran, A 2016, Lithium Ion Batteries in Electric Drive Vehicles, SAE International, Warrendale. Available from: ProQuest Ebook Central. [1 January 2021].
- [14] Mapbox Map Layers in Plotly using Python (2020) [Online], Available at: <https://plotly.com/python/mapbox-layers/> (Accessed 15 December 2020).
- [15] Mapbox HexagonLayer in Pydeck using Python (2020) [Online], Available at: https://deckgl.readthedocs.io/en/latest/gallery/hexagon_layer.html (Accessed 15 December 2020).
- [16] ICCT, the surge of electric vehicles in United States cities (2019) [Online], Available at: The surge of electric vehicles in United States cities | International Council on Clean Transportation (theicct.org) (Accessed 1 January 2021).
- [17] Python's visualisation library Seaborn's Barplot script (2020) [Online], Available at: <https://seaborn.pydata.org/generated/seaborn.barplot.html> (Accessed 15 December 2020).
- [18] Business Insider, Tesla has a huge lead over other automakers. See how the electric-car company's brand became so mighty (2020) [Online], Available at: <https://www.businessinsider.com/why-tesla-has-a-huge-lead-over-competitors-elon-musk-2020-7?r=US&IR=T> (Accessed 1 January 2021).
- [19] Python's visualisation library Matplotlib.pyplot's Barplot script (2020) [Online], Available at: <https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/> (Accessed 15 December 2020).
- [20] Tesla's Model 3 Specifications (2020) [Online], Available at: https://www.tesla.com/en_gb/model3 (Accessed 1 January 2021).
- [21] The Guardian Joe Biden executive orders will reverse Trump on climate, Iran, Covid and more (2021) [Online], Available at: <https://www.theguardian.com/us-news/2021/jan/17/joe-biden-executive-orders-trump-climate-iran-covid> (Accessed 1 January 2021).
- [22] Python's visualisation library Plotly's Map Subplot script (2020) [Online], Available at: <https://plotly.com/python/map-subplots-and-small-multiples/> (Accessed 15 December 2020).