**Detecting Urban Mobility Behaviors using Machine Learning**

Submitted by:

Team Project Number:

Dhanush Dinesh

Priyank Arya

Vinay Shanmukh Aradhya

Zhe Meng

Advisor:

Dr. Nektaria Tryfona

April 20, 2021

Submitted in fulfillment of the requirements for design project

**Bradley Department Electrical and Computer Engineering**

**Virginia Tech**

**Table of contents**

**Urban Mobility trends…………………………………………………………………….……3**

*Abstract .............................................................................................................................3*

**1.0 Introduction ................................................................................................................3**

*1.1 Questions and views ....................................................................................................4*

**2.0 Literature Survey …………………...............................................................................5**

**3.0 Design Method..............................................................................................................6**

*3.1 Clustering…………..........................................................................................................6*

*3.3 Data ................................................................................................................................6*

*3.3 Experiment results ..........................................................................................................7*

*3.4 Visualization.....................................................................................................................8*

**4. Analysis ............................................................................................................................**

*4.1 Results and Contributions ................................................................................................*

**Literature and Source list ...................................................................................................**

**1.0 Introduction:**

This report discusses our key messages and findings from the analysis that we have done for Urban Mobility. Access to urban services is in higher demand in nations like the United States (89 percent of the US population is predicted to live in urban areas by 2050), putting additional strain on our public transportation networks. Based on this pattern, it's evident that a restructuring of urban transportation is required to counteract the expanding population, and we'll show you how to do it. Our society's ability to operate depends on our ability to move about in cities. It serves as a portal to our daily life, allowing us to access housing, employment, and city services. And, while online services have decreased part of the reliance on transportation in the urban mobility scene, they haven't totally eliminated it—rather, they have pushed it to adapt to new needs.

Urban mobility helps to facilitate population expansion by allowing growing communities to access services in a timely and efficient way. Traffic congestion and increased demand for public transportation are two of the most significant consequences of population growth, both of which can be alleviated by changing the current state of urban mobility and moving toward new modes of transportation that go beyond traditional modes like private car ownership and public transportation. Urban growth necessitates the development of sustainable and efficient mobility alternatives and services that evolve in tandem with society. And without a Sustainable Urban Mobility Plan, none of this will be possible.

When travel restrictions were lifted, people were hesitant to travel because of the huge toll that was seen on the lives of the people during the pandemic. Work from home was implemented by most of the Tech Companies which significantly brought a huge change in the mobility pattern. People also became conscious of the spread of the virus from the public space, hence they refrained from the public transportation system. Diseases cause a change in the lives of people.

The Government, especially the transport department of every country faced a huge loss during this time. It’s high time that the Government learns from this situation and makes the best of it.

While the epidemic has afflicted people of all ages, the elderly and children have been significantly more likely to experience the most serious health repercussions. So, the older people and children were refrained from traveling. The outcomes also revealed that during pandemics, travel distances tend to shorten, and journeys become less frequent. Because the majority of the respondents were traveling primarily for shopping during the pandemic, social distance will be necessary for an extended period of time.

**1.1 Question and views:**

We faced the Spanish flu in 1920 and we came across a similar situation almost a century later. Since we are technologically advanced, we were able to figure out a solution in a short span of time. The Covid-19 has shown the world to reassess one’s priorities, highlighting the importance of hygiene and health. With this project, we have tried to answer some of the intriguing questions with respect to the changes that were witnessed post the Covid-19 era. We have tried to understand the impact of Covid-19 on high functional areas such as Hospitals and Schools, analyze the number of passengers of which subway lanes and bus routes were affected, the change in the mobility pattern of the demographic group with respect to Washington DC, and to analyze if the changes witnessed post the pandemic are temporary or permanent.

**2.0 Literature Survey:**

Much recent research focuses on traffic and mobility patterns during the COVID period. Applying twelve scenarios, **Advani, M, Sharma, N and Dhyani, R (2021)** [1] studied Delhi's mobility changes and the demand with non-motorized transportation. The study shows that compared to the pre-COIVD period, unlocking level 3 has a vehicle kilometer traveled (VKT) reduction of 19% in Motorized Two-Wheelers (MTWs), 5% in Cars and 49% in Buses. Private vehicles are4 more preferred because of the infection concern. The increase in bicycle trips has been estimated to be 5.88 million for the post-lockdown Scenario compared to 1.1 million trips estimated for the pre-COVID Scenario. As a result of the significant modal shift from motorized to non-motorized, a decrease in both vehicular emission and road accidents is observed **(Alfredo Aloi et al. 2020; Abdullah, M et al. 2021) [2]**.

As a powerful tool, machine learning is often used in COVID case tracing and urban mobility modeling separately. Combining clustering and the feature selection techniques, **Khmaissia F et al. (2020)** determined patterns of ZIP code-level increase in the number of new COVID-19 cases in megacities like NYC [3]. And by using the machine learning technique, **Kuo, C.-P and Fu, J. (2020)** indicated that, compared with the Phase I re-opening, a 1-week and a 2-week lockdown could reduce 4% – 29% and 15% –55% infections, respectively, in the future week, while the 2-week Phase III re-opening could increase 16% – 80% infections [4]. Along with ML applications in virus tracing, in research about Urban Mobility, **Song H.Y and You D (2018)** established a method using Clustering techniques (DBSCAN and GMM) to identify and analyze urban mobility models based on real taxi transportation data [5]. In another study focusing on regional mobility patterns, combined with the nonnegative tensor factorization, the clustering technique (fuzzy C-means) is also used to provide more meaningful region division and higher interpretability of the extracted data **(Qi G et al. 2019) [6]**. Moreover, **Vidovic K, Mandzuka S and Brcic D (2017)** proposed an adaptive neuro-fuzzy inference system (ANFIS) to build an urban mobility index, enabling a new approach for urban mobility assessment based on real domain expert's expertise [7].

Compared with the large quantity of research solely about ML and urban mobility, a few pioneer studies have tried to apply the ML method to analyze the impact of urban mobility during this worldwide pandemic period. Using the biclustering algorithm to discover different traffic patterns, **Aparicio, J. T, Arsenio, E, and Henriques, R (2021)** introduced a dynamic assessment of the effects of COVID-19 on public transport use [8]. Besides, **Paiva, S et al. (2022)** used the K-means clustering method to analyze how specific age groups' mobility changed under the COVID period [9]. And as one of the results, they found that the mobility in retail and recreation areas is reduced compared to the pre-pandemic period.

**3.0 Design method:**

**3.1 Clustering:**

While working with unstructured and unclassified datasets, clustering techniques are often used to find and group similar entities together so that insights can be derived from these groups. We used clustering as we wanted to visualize how the data points are distributed across the space and group similar data points together so that we can analyze and profile the underlying attributes of each group and derive insights in the mobility pattern changes across the socio-demographics.

**Different methods of clustering**

Few of the popular clustering techniques are KMeans, Affinity Propagation, Hierarchical or Agglomerative clustering, and DBSCAN.

KMeans clustering is used for partitioning an N-dimensional population into k clusters, where the value of k is to be specified prior to clustering. It is a widely used technique that makes clusters based on geometric distances between points. The clusters are grouped around centroids, causing them to be globular which is one of its major drawbacks also. So if the underlying clusters are not globular then K-Means produces poor results.

Affinity Propagation makes clusters based on the graph distances between the points leading to smaller uneven clusters. The user doesn't have to specify the number of clusters like in KMeans, but it does not produce good results if the underlying clusters are non globular. Also, it is difficult to scale for large datasets as it is computationally expensive.

Hierarchical or Agglomerative clustering generates a hierarchy of clusters. It is good for non globular clusters and scales well to larger datasets but just like K Means, the user must specify the number of clusters prior to clustering.

DBSCAN is a density based algorithm that makes clusters for dense regions of points. It is great at separating clusters of high density clusters versus clusters of low density and it performs well with arbitrary shaped clusters. This fits perfectly for our dataset as the data will be dense at some places like schools, business districts, shopping malls, while being sparse at other places like suburban areas.

In addition, DBSCAN doesn’t require every point to be assigned to a cluster thereby reducing the noise of the clusters. And based on the parameters of epsilon and minPts, we can classify the points as core point, border point or outlier.

Silhouette Coefficient is a metric which is used to evaluate the goodness of a clustering technique. We implemented the above four clustering techniques on our datasets and then calculated their Silhouette Coefficient. Based on the Silhouette Coefficient, DBSCAN provided the best results and hence we chose this technique.

**DBSCAN**

DBSCAN (density-based spatial clustering of applications with noise) is a well-known machine learning data clustering technique. The DBSCAN technique is used to uncover correlations and structures in data that are difficult to find manually but are meaningful and valuable in predicting patterns and trends. And we have also taken into consideration the C means Clustering method as well. Fuzzy logic principles may be used to cluster multidimensional data by assigning each point a percentage membership in each cluster center ranging from 0 to 100%. When compared to standard hard-threshold clustering, which assigns a clear, accurate label to every point, this can be highly strong [10]. On the basis of the distance between the cluster center and the data point, this method assigns membership to each data point that corresponds to each cluster center.

**3.2 data**

**Raw Data**

There are two kinds of data used for this project: census data and mobility data. The Census data is available on the official website of the U.S. Census Bureau, which provides data relating to the population specifying a designated area. Mobility data includes smartrip data and GeoDS Lab Data[11].

**Smartrip data**

Our hourly research is based on Smart trip data. The data contain the main table providing Entry & Exit information on dates, times and specific station IDs, and a mapping table containing station IDs, names, and geographic information. The smart trip data and mapping data samples are presented in Table 1 and Table2. Smart trip data are composed of CUSTOMER\_CODE, ENTRY\_MSTN\_ID, EXIT\_MSTN\_ID, ENTRY\_DTM and EXIT\_DTM. The CUSTOMER\_CODE stands for the customer taking the specific trip. Together with ENTRY\_MSTN\_ID, EXIT\_MSTN\_ID and the corresponding information in the mapping table, we can get the entrance and exit information of the specific trip. ENTRY\_DTM and EXIT\_DTM stand for the date and time of the trip. For example, the first row of Table1 can be interpreted as one person started a trip at the station Stadium Armory at 0:46 on 5/1/2016 and ended the trip at the station Cheverly at 1:00 on 5/1/2016.

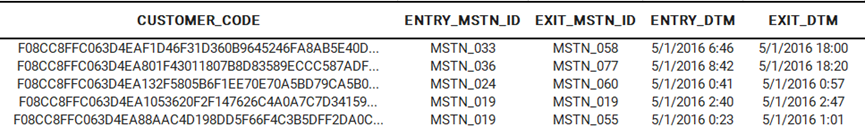


Table 1



Table2

**GeoDS Lab Data**

Research of daily patterns depends on the GeoDS Lab Data[11], which is openly available on Github. This data results from Kang, Y., Gao, S., Liang, Y. Li, M., Rao, J. and Kruse, J. Multiscale from the Geospatial Data Science Lab of UW-Madison. “By analyzing millions of anonymous mobile phone users’ visit trajectories to various places provided by SafeGraph”[11], they estimated “the daily and weekly dynamic origin-to-destination (O-D) population flows”. And the data were “computed, aggregated, and inferred at three geographic scales: census tract, county, and state.” Here, we chose the level of the census tract. The Sample of the initial GeoDS Lab data is presented in Table3. The GeoDS Lab data include geoid\_o, geoid\_d, lng\_o, lat\_o, lng\_d, lat\_d, date, visitor\_flows and pop\_flows. The geoid\_o, lng\_o and lat\_o stand for the geographic information of the origin, while correspondingly, geoid\_d, lng\_d and lat\_d stand for the geographic information of the destination. Vistor\_flows stands for the exact number of moving people reported by the mobile phone. And pop\_flows stands for the inferred population level of dynamic O-D flows[11]. We use pop\_flows rather than vistor\_flows as the source data in our project. For instance, the first row of Table3 can be interpreted as a trip of one visitor starting at the location with geo ID 51510200303 and ending at the location with geo ID 24033802001 on Jan 6th , 2020. And the inferred number of population flows resulting from this trip is 8.

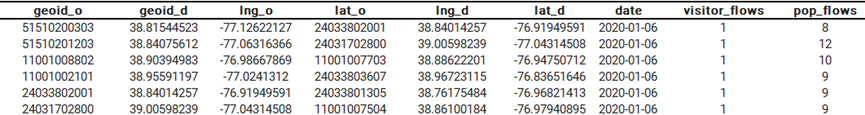


Table3

**Data Preprocessing**

Before implementing the Machine Learning algorithm, Our raw data needed some modifications to fit the requirement of our potential Machine Learning models. Generally, the transformation included three main steps: Filtering, Merging and Calculating for both smart trip and GeoDS Lab Data. Some extra changes were executed according to specific properties of the data sets. Pic 1 shows the workflow of data preprocessing.



Fig 1

**Smartrip data**

Generally speaking, the filtering step for smart trip data was grouping records by date and time, cutting the whole dataset into two pieces for each day. One named date\_morning, which included the data from 5 am to 10 am, and another named date\_evening, had the data from 3 pm to 7 pm, dropping out data from other periods. Then came the merging step. We converted the data sets from customer basis to station basis. Each row of the raw data would be counted in both the corresponding entry station and exit station, added separately in ENTRY\_COUNT and EXIT\_COUNT. After all, we combined the ENTRY\_COUNT and EXIT\_COUNT of each station and then concatenated the station name and geographic information from the mapping table. For example, if Table1 represented the whole data for May 1st, 2016, Table 4 and Table 5 would be created, naming 2016\_5\_1\_morning.csv and 2016\_5\_1\_evening.csv. And Table 6 is a real snippet of the cleaned data.

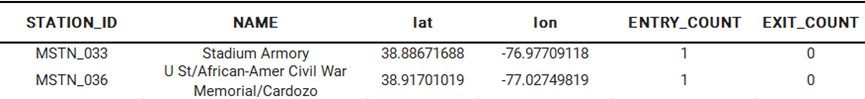


Table 4

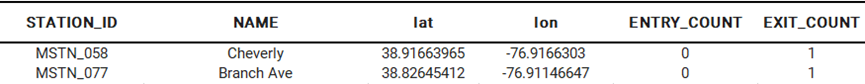


Table 5

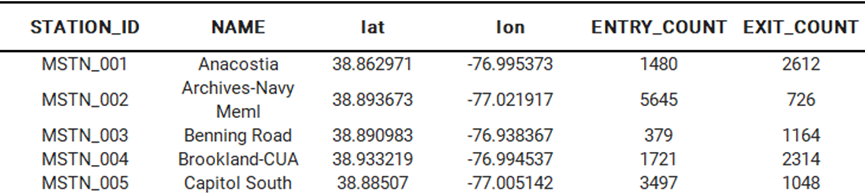


Table 6

**GeoDS Lab Data**

The initial GeoDS Lab Data contained records of the whole nation. We first filtered data sets by our target latitude and longitude, only keeping records having their origin or destination inside NCR. The raw data sets were trip-based, and we should convert them into location-based ones. So, similar to what we did with the smart trip data, we separated each row into the exit and entry parts. And then, we assigned the numbers of flows separately according to the corresponding location, under EXIT\_COUNT and ENTRY\_COUNT. Finally, we added the numbers of EXIT\_COUNT and ENTRY\_COUNT and assigned them to each location. As an instance, if Table 3 is the data set of Jan 6th , 2020. We would create a table like Table 7 below and name it 2020\_01\_06.csv. And Table 8 is a real snippet of the cleaned data.

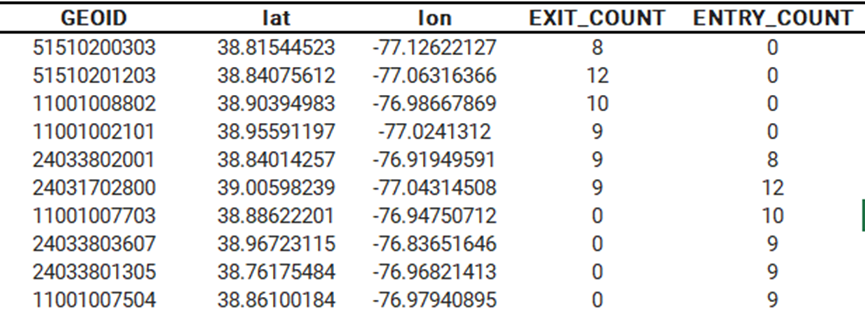


Table 7

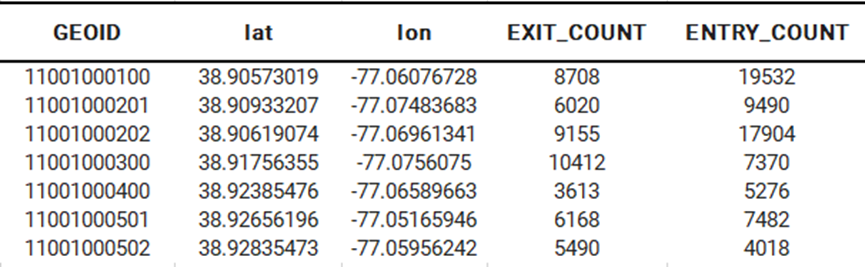


Table 8

Cleaned data

After preprocessing, the initial cleaned data consist of GEOID, lat, lon, EXIT\_COUNT and ENTRY\_COUNT, as shown in Table 8. The Data files are daily based created, named after the occurring date. GEOID, is the identifier of NCR, census tract level. And lat &lon stands by the latitude and longitude corresponding to the GEOID. EXIT\_COUNT stands for the estimated leaving population counts, while ENTRY\_COUNT stands for the estimated entering population counts. For example, assuming the data file is named 2020\_01\_20.csv, the first row of Table 8 indicates that on that day, GEOID 11001000100 had 8708 people come in and 19532 people left out.

**Evaluation:**

|  | Clustering method | Silhouette Coefficient |
| --- | --- | --- |
| 1. | Affinity Propagation | 0.67 |
| 2. | Agglomerative | 0.56 |
| 3. | K means | 0.53 |
| 4. | OPTICS | 0.24 |
| 5. | DBSCAN | 0.16 |

Table1: Clustering evaluation using silhouette score

We observe we have calculated the silhouette score of each of the clustering methods in Table.1. The lowest score signifies a better fit of the clustering algorithm. The best clustering technique we observed for our data was the DBSAN which had a silhouette score of 0.16. The second best clustering technique was the OPTICS method. Hence we chose the DBSCAN for our implementation of clustering.

**4. Visualizations:**

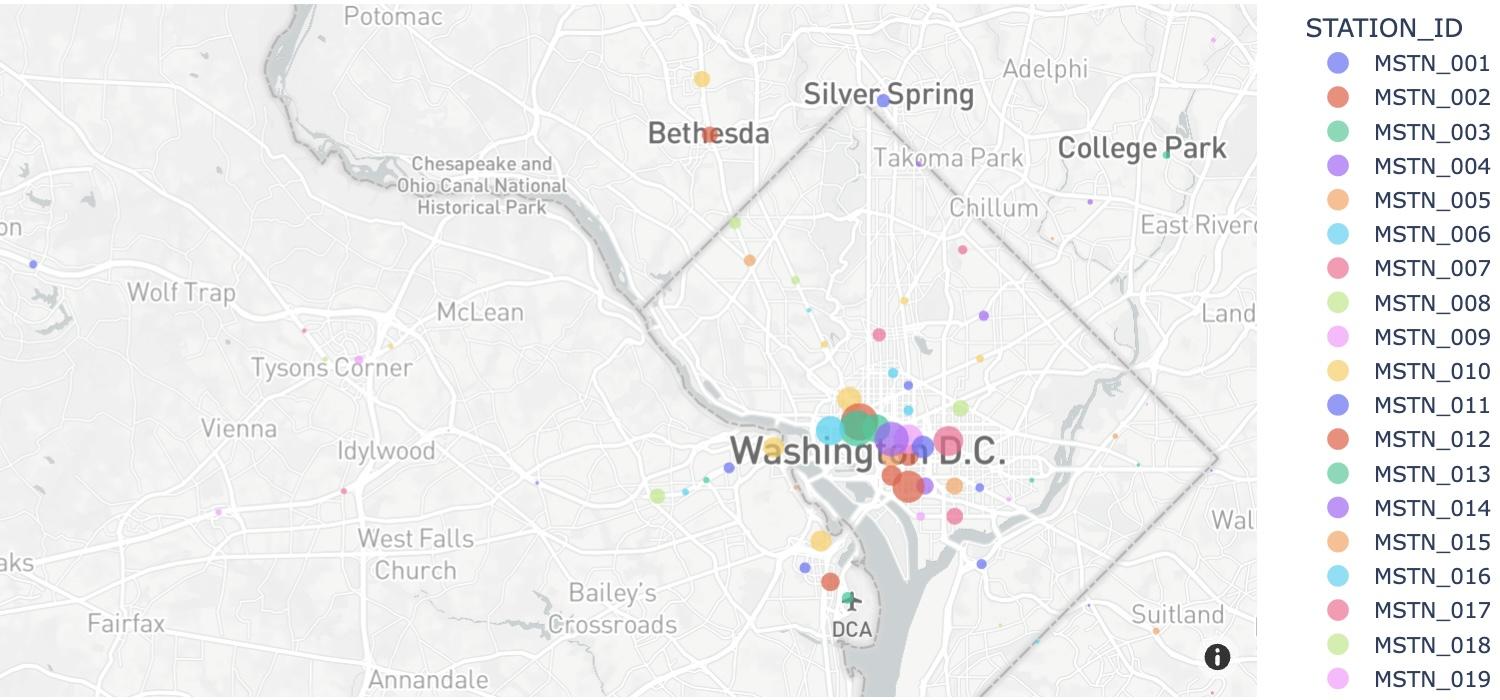
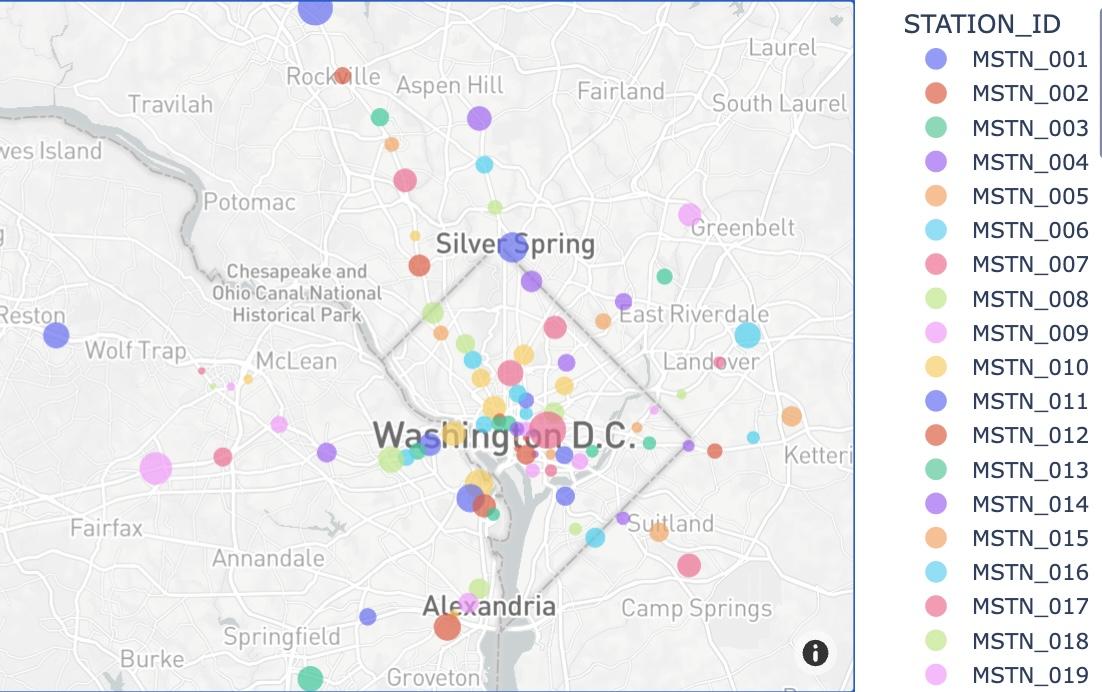


Fig1: Morning Entry, Smart trip data Fig2: Morning Exit, Smart trip data

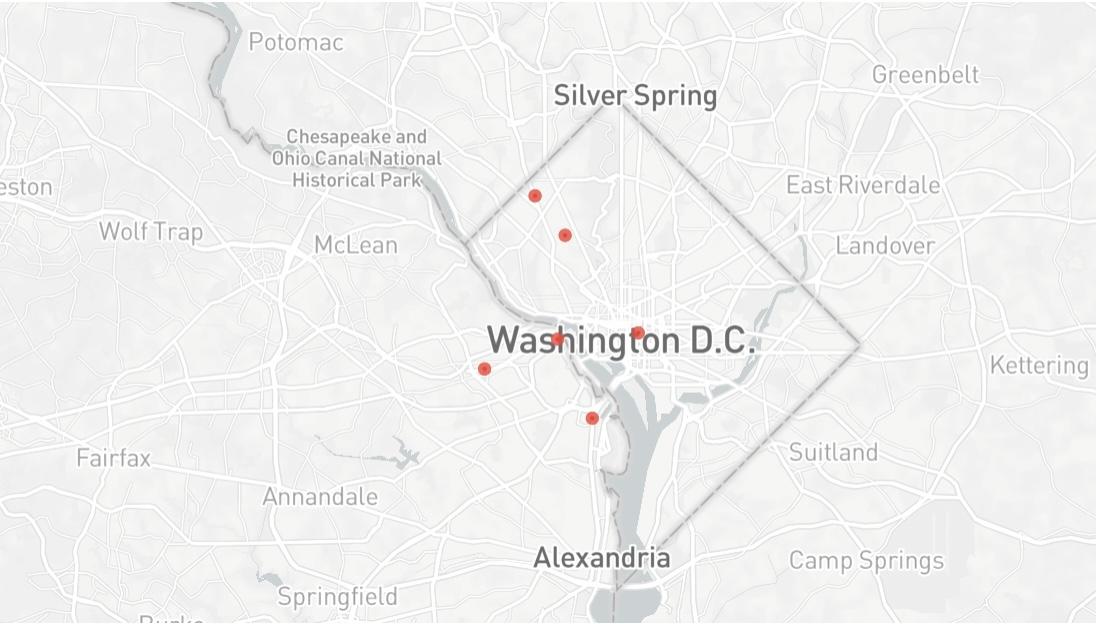
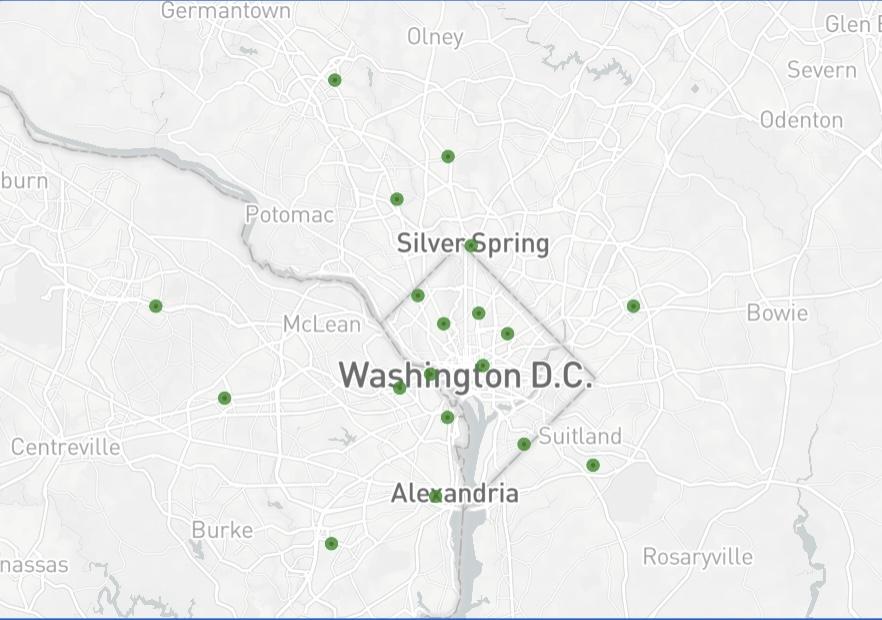


Fig3: Morning Entry, Smart trip data cluster data with 19 clusters

Fig4: Morning Exit, Smart trip data cluster data with 6 clusters

The distribution of the clusters corresponds to the DC metro lines as shown in Fig1. The data that was collected for the Entry points on the Morning of May 2, 2016 we can see that we have obtained 19 clusters using the DBScan technique as shown in the Fig3. We were able to obtain 6 clusters using the DBScan technique for the exit points of travelers on May 2, 2016. These clusters are centered around the US Capitol region, Washington National Cathedral and Ronald Reagan International airport. These indicate that the majority of the people’s exit has centered around the attraction centers in the DC area. This depicts the apt representation of the movement of the people with respect to a federal holiday when there are special events organized for the general population around the major attraction centers in DC.

We can say that the daily flow of the people are mostly from suburbs to the city center in the morning and from city center to the suburbs in the evening. We can observe that in the graph Fig1 and 2. The interesting fact that there are more clusters in the morning with more entrances signifies the wide distribution of the points. Due to the ensits being more focused in the center of the city we can observe less clusters.

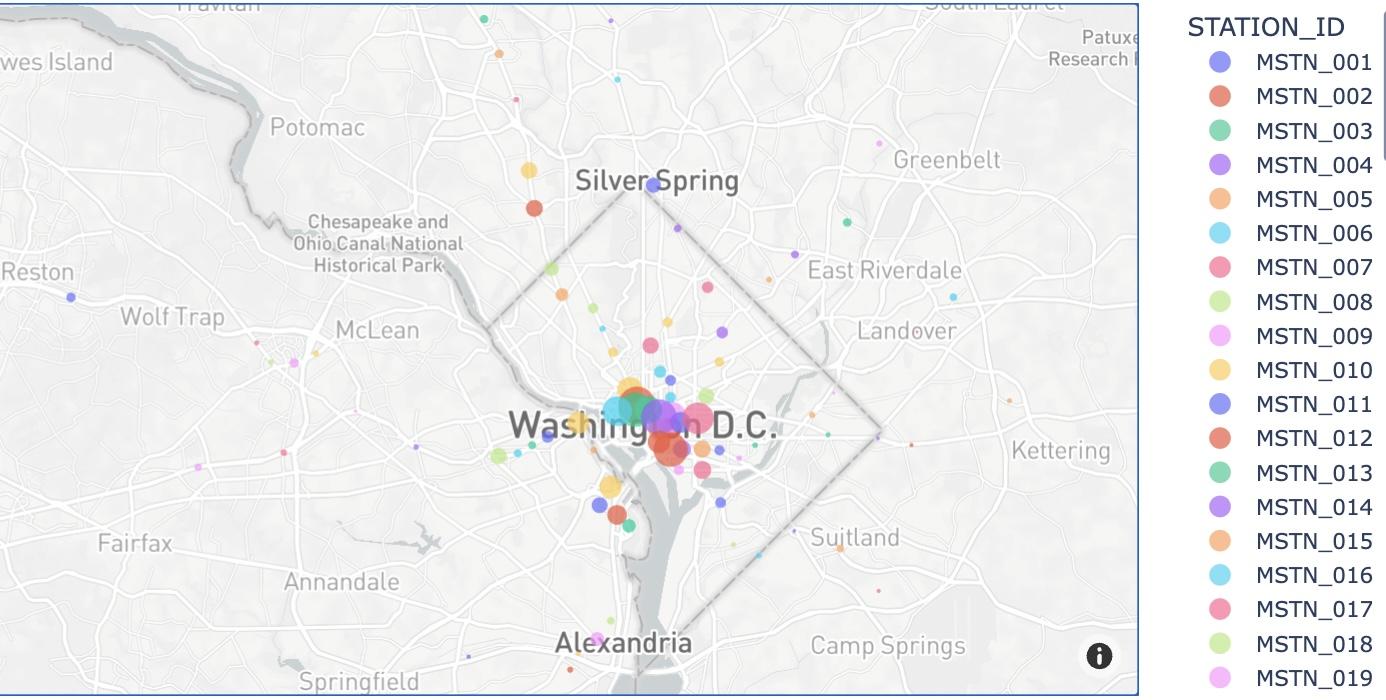
 

Fig.5: Evening Entry, Smart trip data Fig.6: Evening Exit, Smart trip data

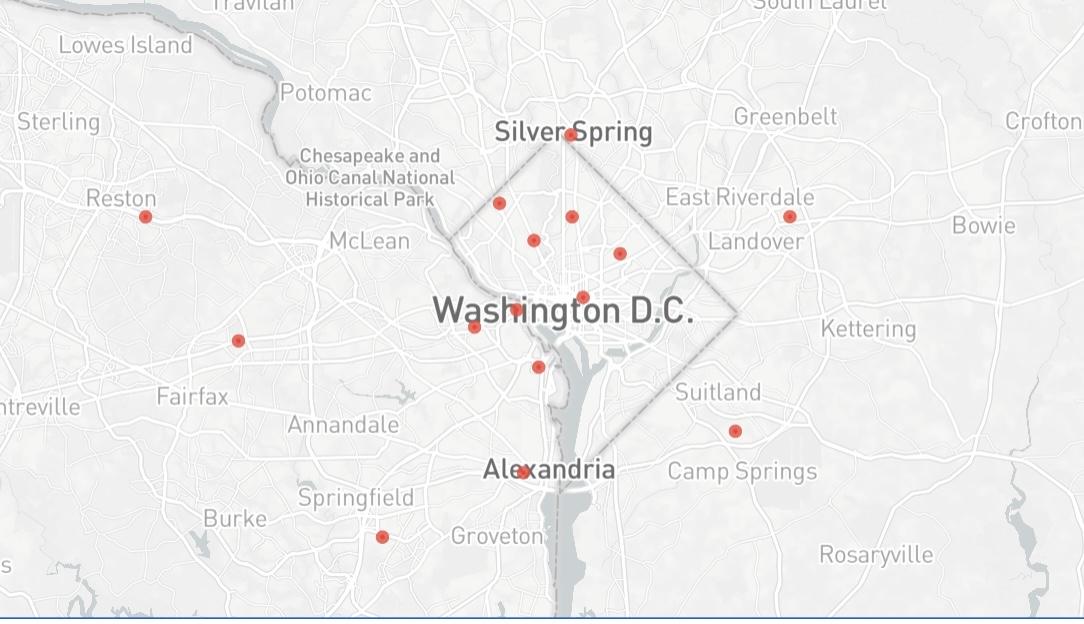
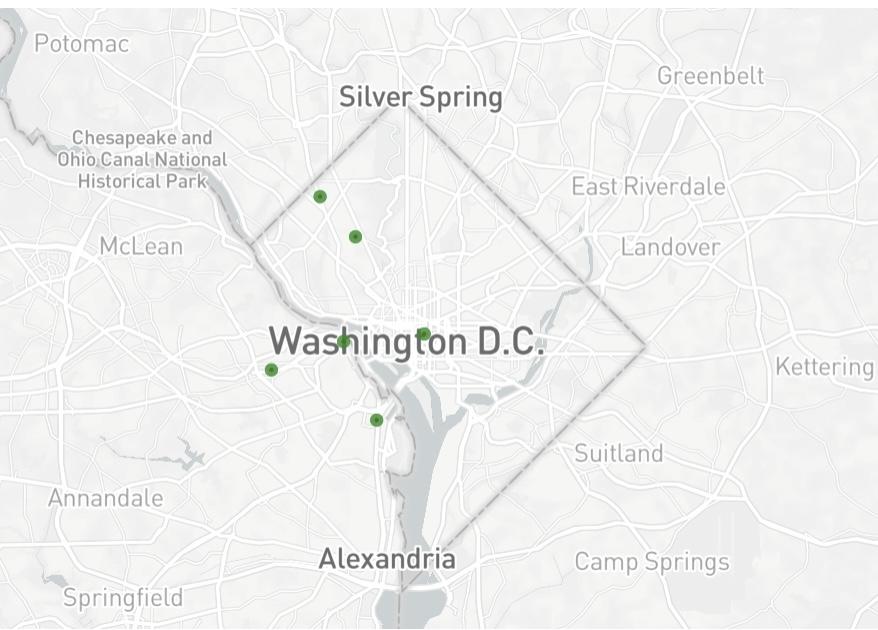


Fig.7: Evening Entry, Smart trip data Fig.8: Evening Exit, Smart trip data cluster data cluster data with 6 clusters with 16 clusters

We observed 6 clusters in the central DC area for May 2nd 2016 as shown in Fig.7. The entry points of people for the evening of Labor day. We can observe the size of clusters that are comparatively larger than the exit points of Morning indicating that there is a larger number of people who are exiting from the central region of DC and are moving to their respective places. This explains the fact that there are comparatively a larger number of people who were present near the major monumental landmarks of DC that includes the White House, United States Capitol, Lincoln memorial and so on.

This exit cluster for the evening of May 2nd 2016 shows the exit points of people who took the metro. This indicates that the people’s exits are spread throughout the spread of the metro lines. The major clusters are spread until East Riverdale, Sultland, Alexandria, Fairfax, Gaithersburg which means that people have traveled to these places at the end of the day. Major clusters are seen in the central DC area because of the day being a federal holiday and the inflow of people to the capitol could be seen in the evening as well.

We can compare the Fig.3 the clustering data of Morning Entry with 19 clusters and Fig.8 the clustering data of evening entry with 16 clusters here we observe that the clusters are similar but there is a decrease in the number of clusters in the evening all the 3 clusters which were missing are observed to be near Silver Spring. This suggests that either the people have not traveled back or they have used other modes of transportation.

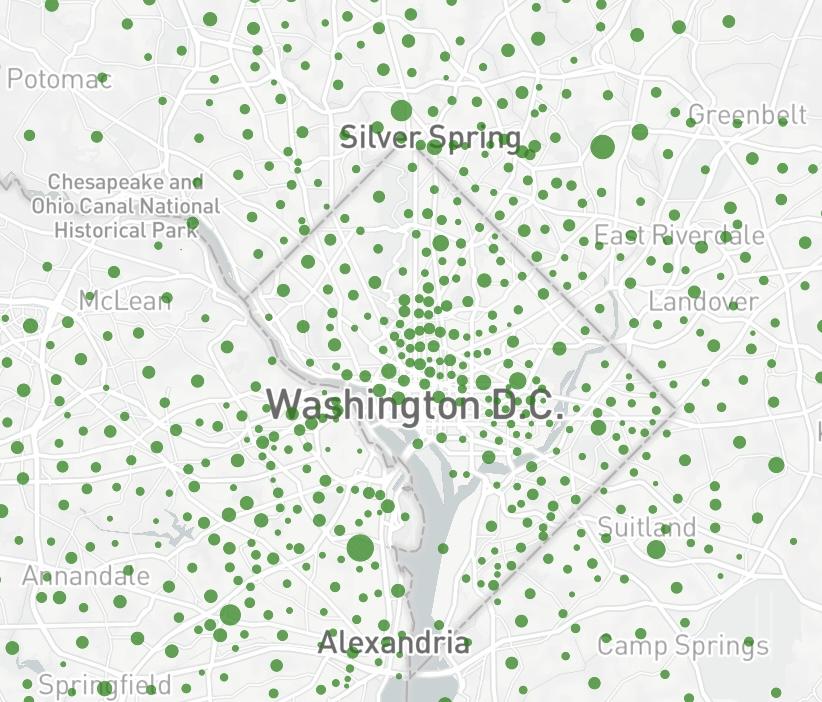
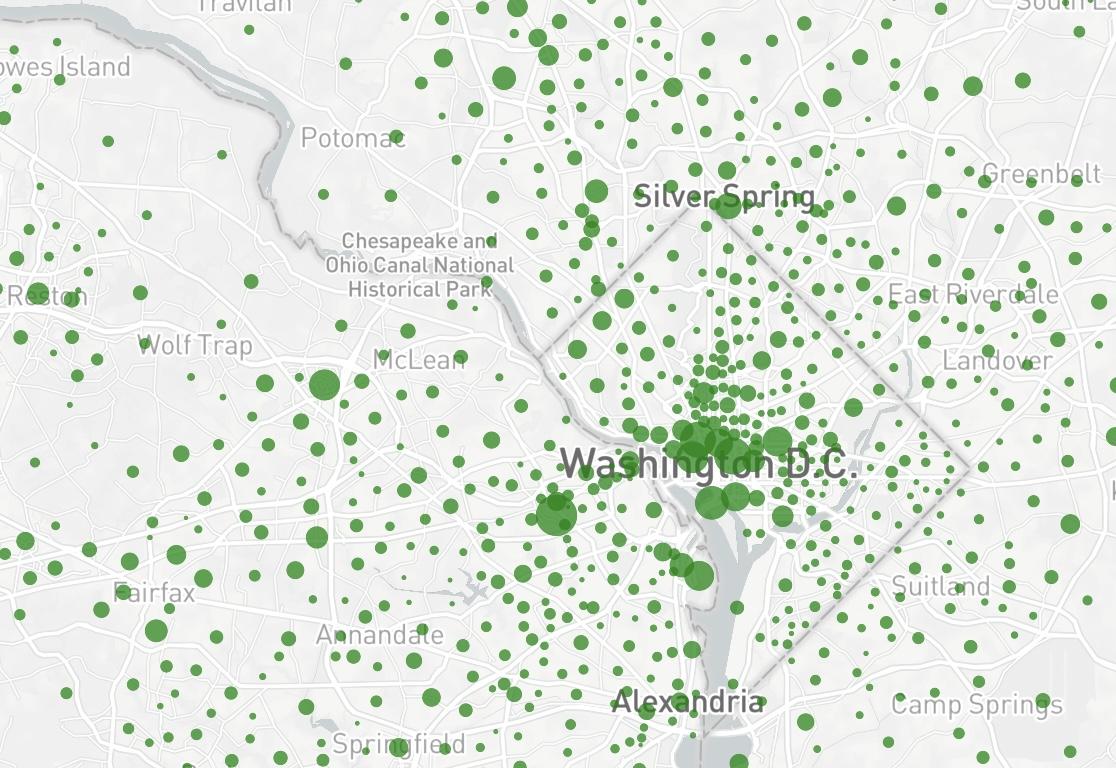


Fig9: Jan 22, 2020 Entry, GeoDS Lab Data Fig10: Jan 27, 2021 Entry, GeoDS Lab Data

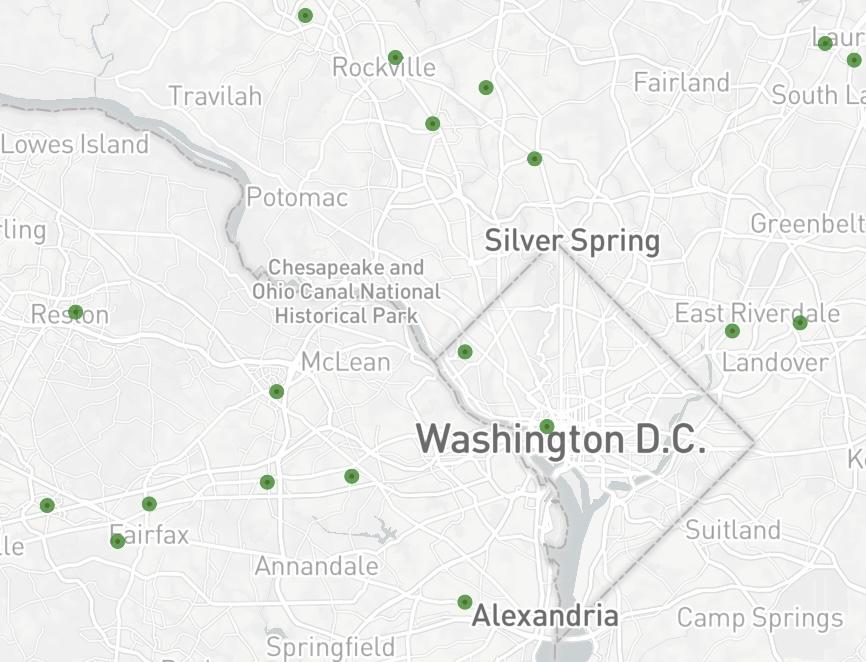
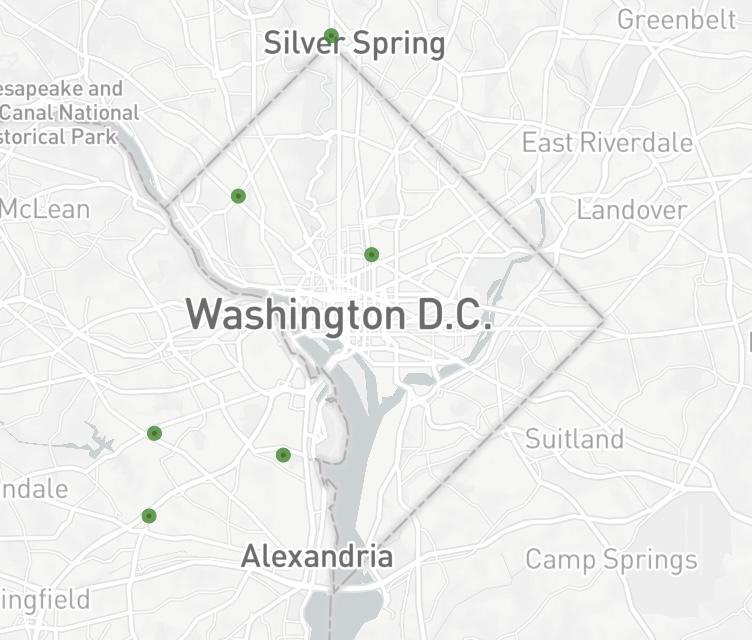
 

Fig11: Jan 22, 2020, Entry clusters (16) Fig12: Jan 27, 2021, Entry clusters (6)

We can observe that in Fig9 the points are much bigger than that of the points in Fig10 which is of 2021. It can be inferred that the traffic flow was reduced considerably in 2021 after the pandemic.

In Fig 11 we have 16 clusters formed that spread throughout the DC region. Jan 6 2020 was a Monday and the major clusters are seen in the Capitol region, IAD and RR airports. Jan 2021 was the time when Covid-19 was at the peak in the US and people were experiencing the harsh effects of the virus. These clusters could mean that people are traveling without any restriction due to the virus. And the significant clusters in the capitol could mean that there were more people who were visiting the DC region.

These are clusters seen for the post covid times where the spread of the clusters have seen to be evenly spread throughout the region. But compared to the Clusters seen in the Fig12, it is evident that visitors/ travelers are reduced because of the pandemic. Looking closely at the size of clusters, Congressional cemetery, Jesuit community, Oak hill cemetery have a lot of hotspots.

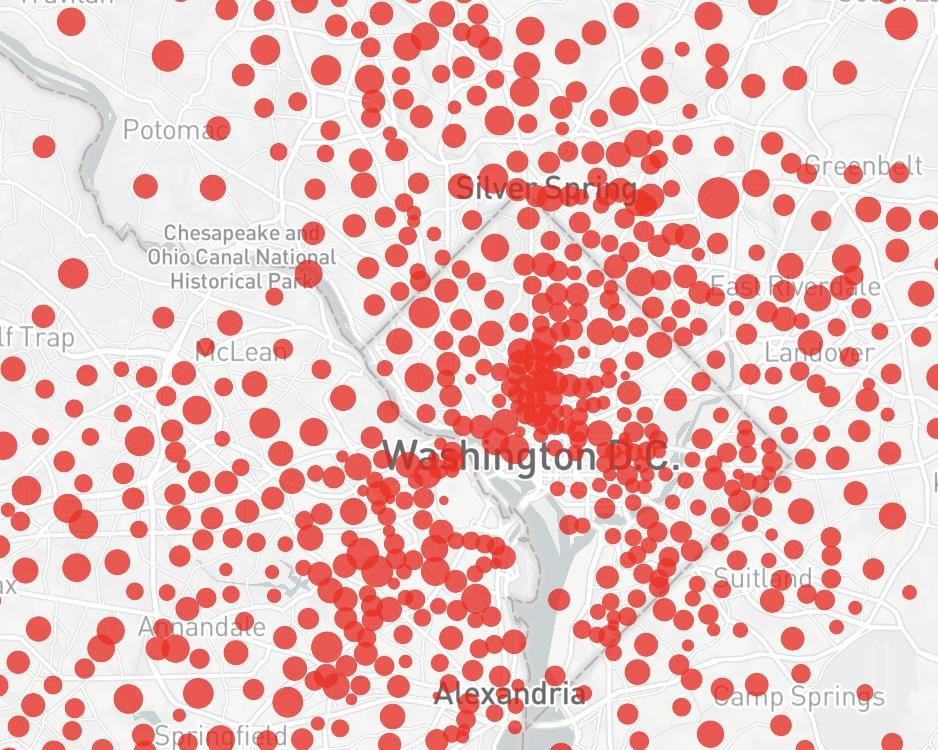
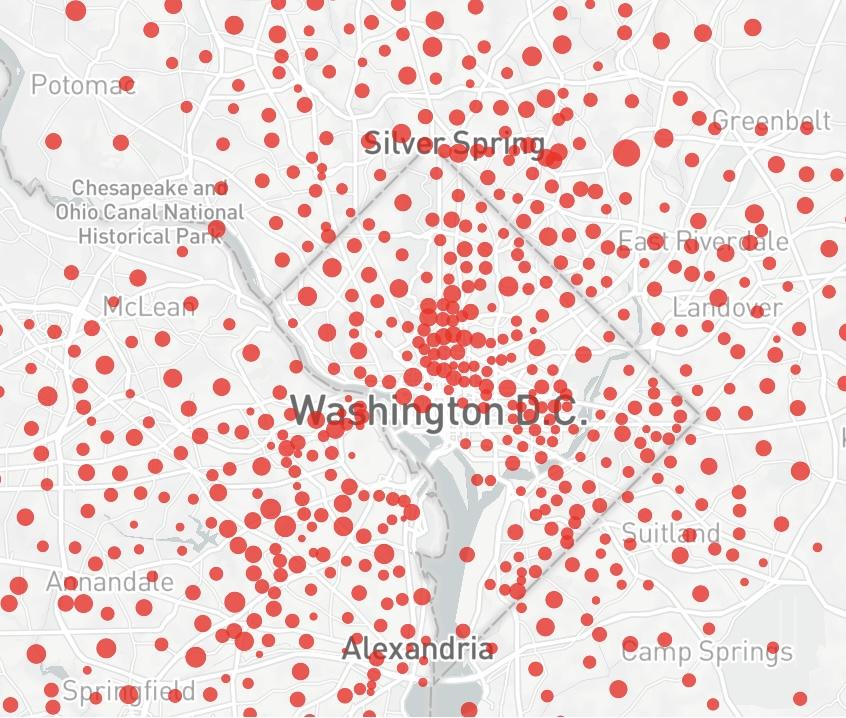
 

Fig13: Jan 22, 2020 Exit, GeoDS Lab Data Fig14: Jan 27, 2021 Exit, GeoDS Lab Data

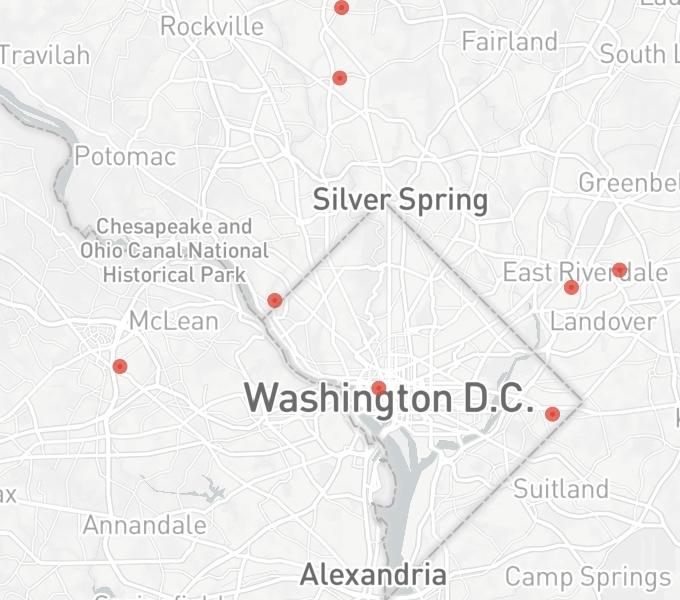
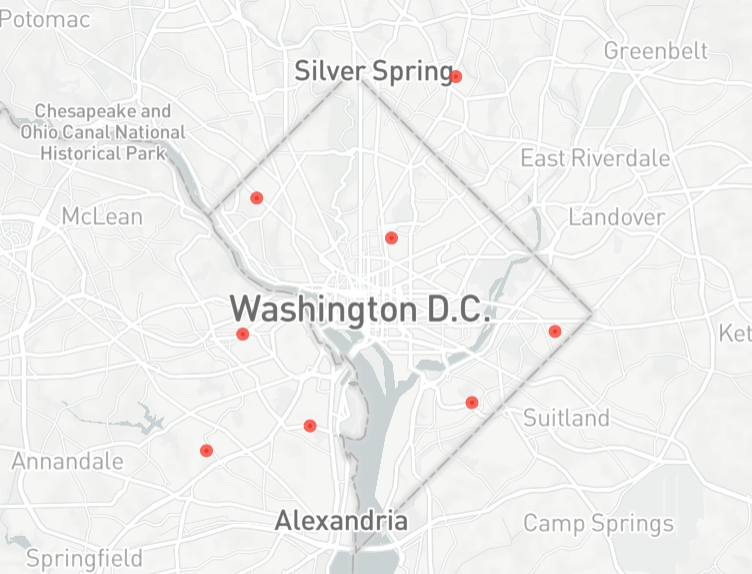
 

Fig11: Jan 22, 2020, Exit clusters (10) Fig12: Jan 27, 2021, Exit clusters (7)

In Fig 13 we see that the plot is similar to the entry as shown in the fig9 as both represent precovid data. We see that the number reduces as shown in Fig 14. We also observe that on Jan 22, 2020 we can observe that there are around 10 clusters. On the other hand in Fig12 we can see that there are around 7 clusters which are evenly spread out which indicates people are not traveling long distances after covid.

**4. Analysis**

***4.1 Results and Contributions***

We have considered all the points we obtained in the clustering technique to figure out what is the impact of covid. We have divided all the points into three main categories as shown in Fig 13

* All the points in blue are cluster points where the traffic flow was less compared the percovid
* All the points in orange are cluster points where the traffic flow has increased after covid
* All the points in orange are cluster points where the traffic flow has increased.

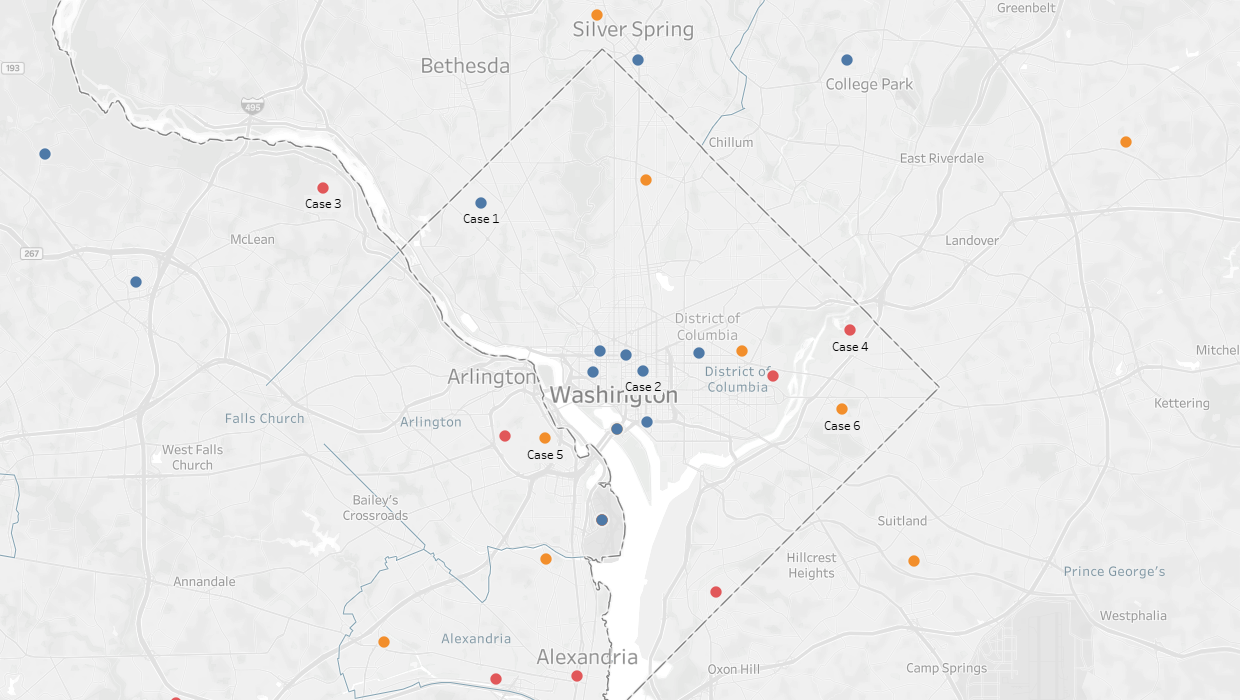


Fig 13. All the clustering points with color coding of their categories

**4.1.1: All the points in blue are cluster points where the traffic flow was less compared the percovid**

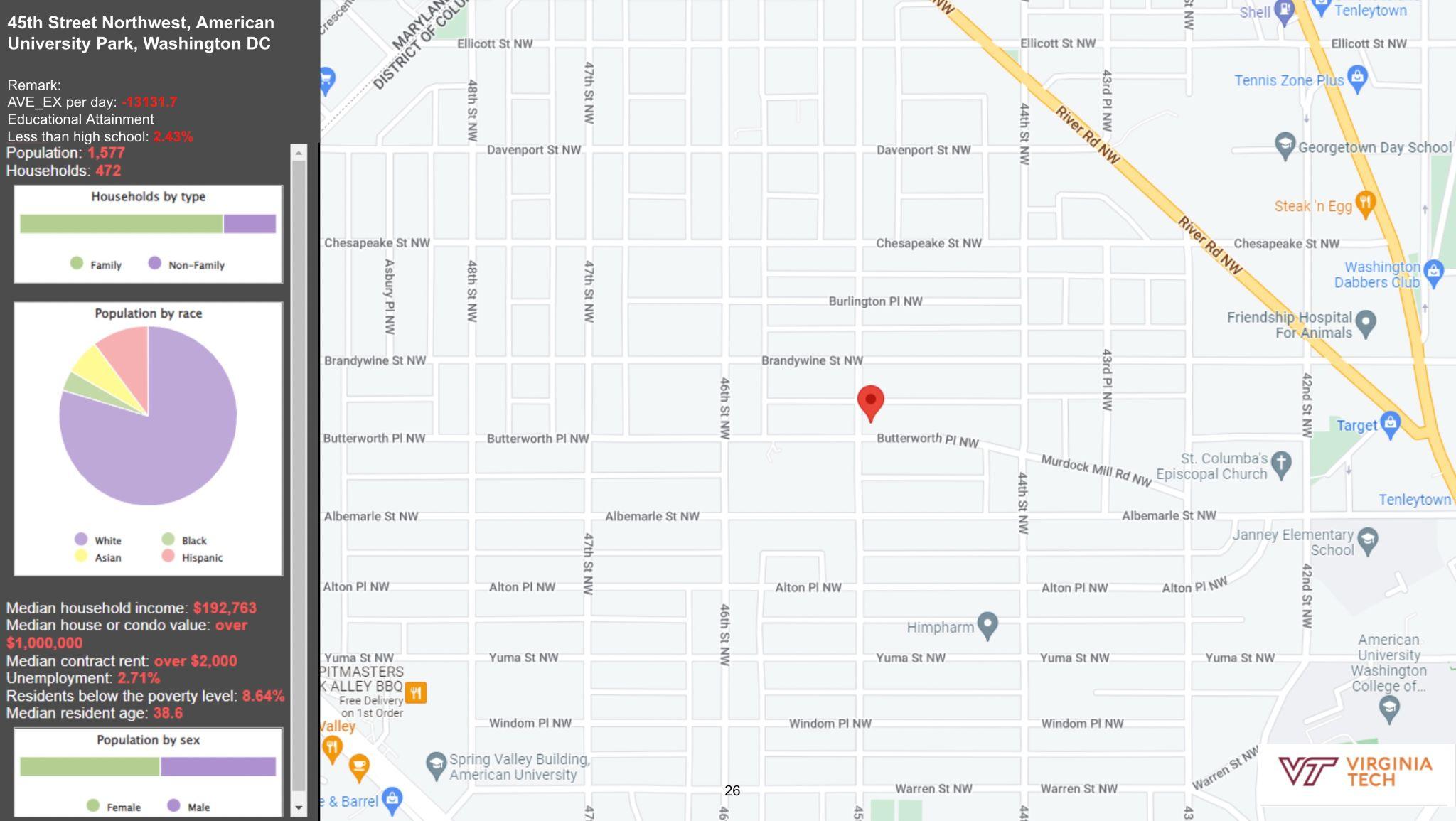


Fig 17: Infographic of American University Park, DC

In Fig 14 the median household income is around 192,763$. We also observe the area is filled with white people. The unemployment rate is very less around 2.7%, with the average age of 40 years and high school dropout less than 3%. We also observed 13 thousand less people traveling in this area, which is a drastic reduction.

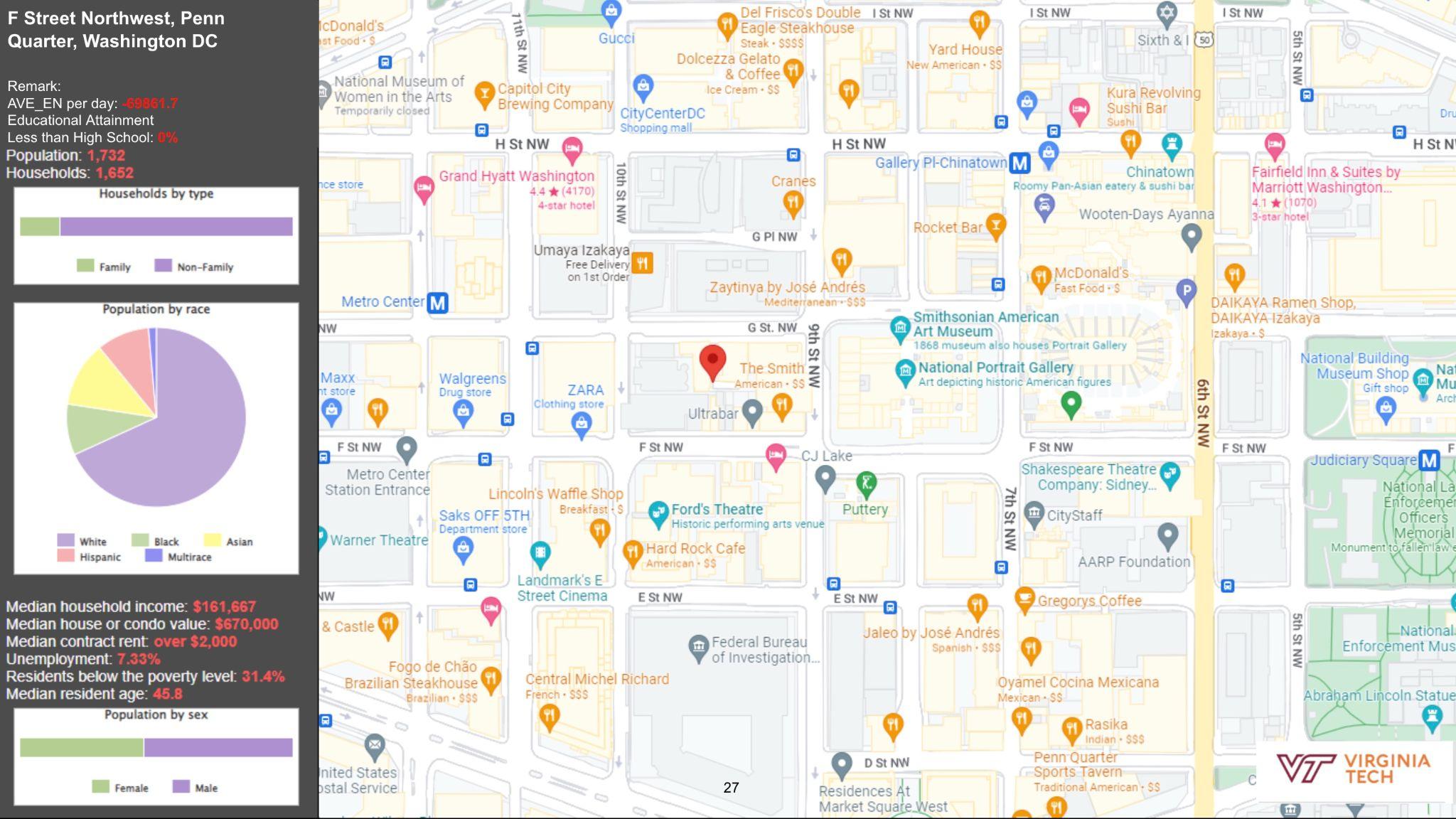
****

Fig 14: Infographic of Penn Quarter, DC

Penn Quarter is predominantly filled with offices, museums and threats as shown in the fig 14. The median household income is around 161,667$, well educated people living with no person who has not finished high school. Most of the demographics is filled with white people with a mix of few asians and black people. A very low unemployment of 7.3% where the nation average is 10.4%. We saw a drop of 69 thousand people traveling to Penn Quarter.

Overall the places where there was a massive reduction of people traveling were universities such as UMD, George Washington University Alumni House, American university park,The George Washington University, DC saw the most decrease. John F. Kennedy Center for the Performing Arts, DC and Thomas Jefferson memorial saw a major decrease in visits which indicates that in 2020 there were a lot less tourists visiting DC. The wealthiest parts of the city such as Wesley Heights saw the most decrease, which has a median household income of 250,000 and all the people living are highly educated. We also saw that Union station,DC have a massive decrease in foot fall.

Most of the neighborhoods we analyzed had around 64-100% whites residents with an average household income in these areas were around 133K$, average unemployment rate was around 4.8% and mean age was 37 years.

**4.1.2: All the points in orange are cluster points where the traffic flow has increased after covid**

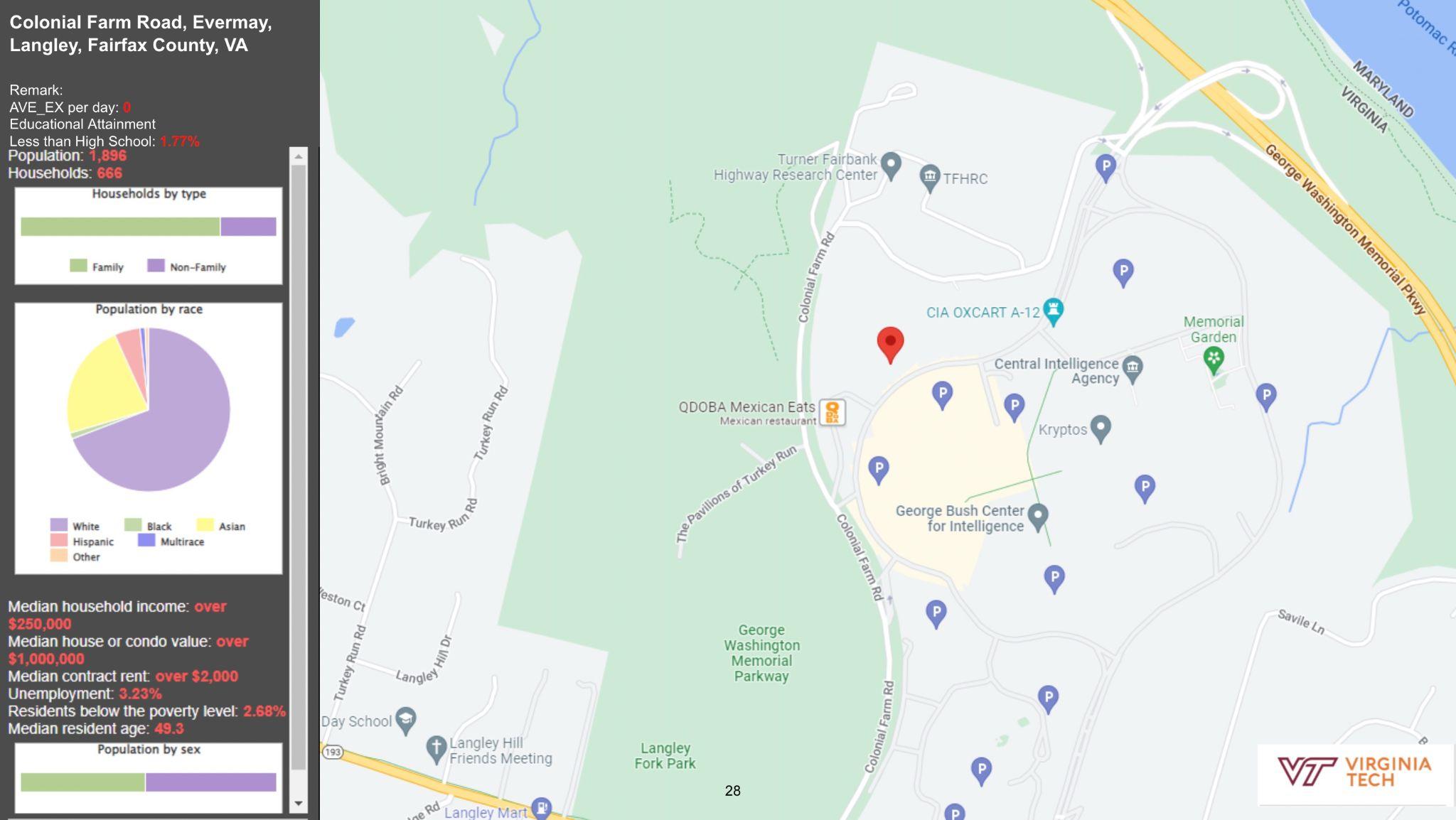


Fig 14: Infographic of CIA HQ Fairfax, VA

In figure 14 the cluster point was CIA Headquarters Fairfax,VA. Federal office which worked through the pandemic saw absolutely no change with respect to the precovid state which explains the no change in traffic flow. We observe a very well educated population living in this area, with mostly whites and asians living, the median income here is over 250,000$. The unemployment rate is a meager 3%. Pentagon City, VA where a lot of federal employees stay saw a slight dip but still did not have a huge change similar to the CIA HQ..

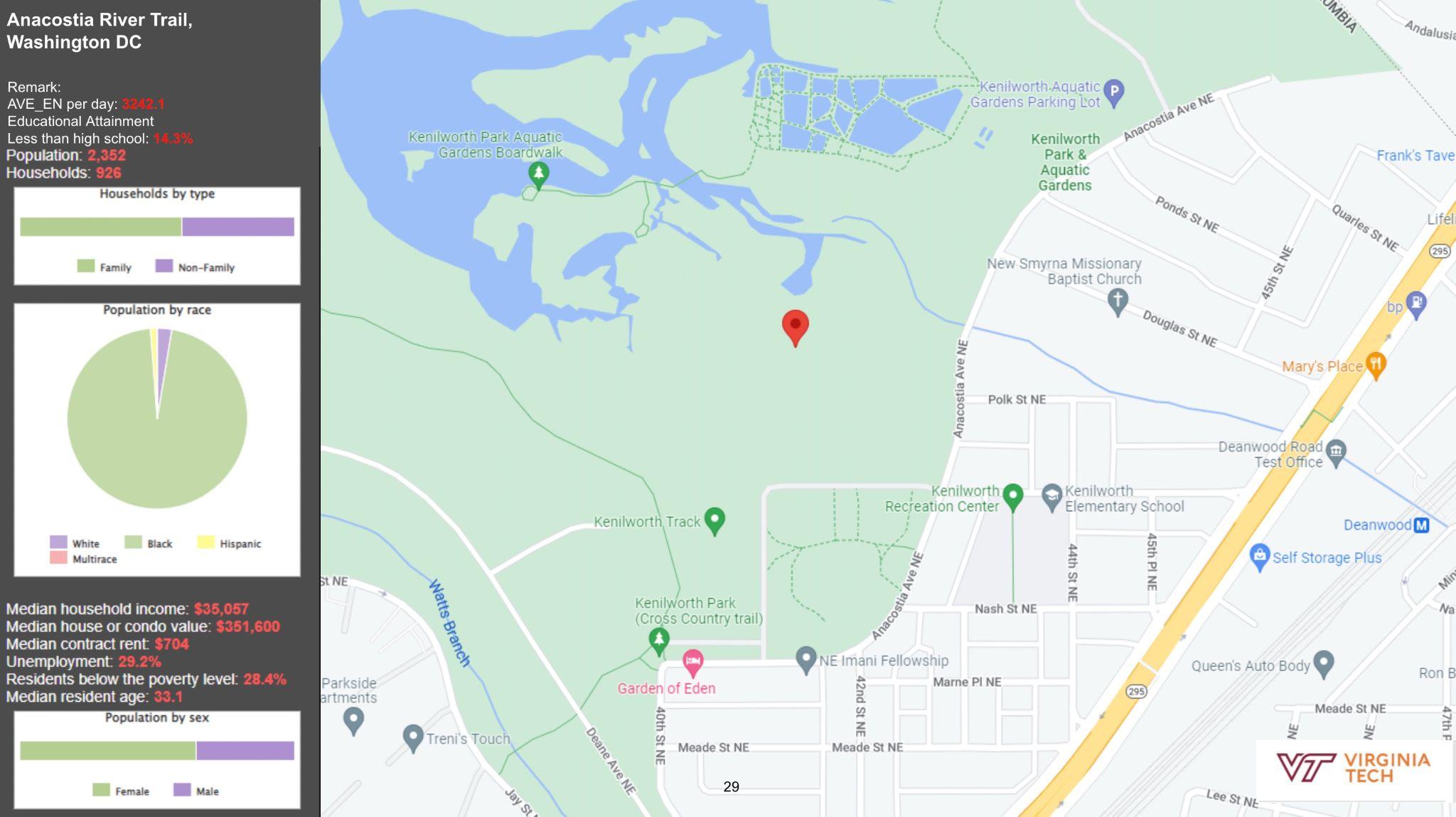
****

Fig 14: Infographic of National capital park,DC (Change)

We observed the area around a grocery/ shopping center where we found absolutely no change

National capital park where we observe there was a slight dip in numbers but as this is an open air center people still felt safe to visit parks. Silver Spring, MD saw the least change in traffic flow. We observe that the people living in these areas are mostly Hispanic around 65.5% with Whites and Asians covering the rest, average household income is around 94,375$, with average unemployment rate of 8.9%. The percentage of people not completing high school was 8.8%.

**4.1.3: All the points in orange are cluster points where the traffic flow has increased.**

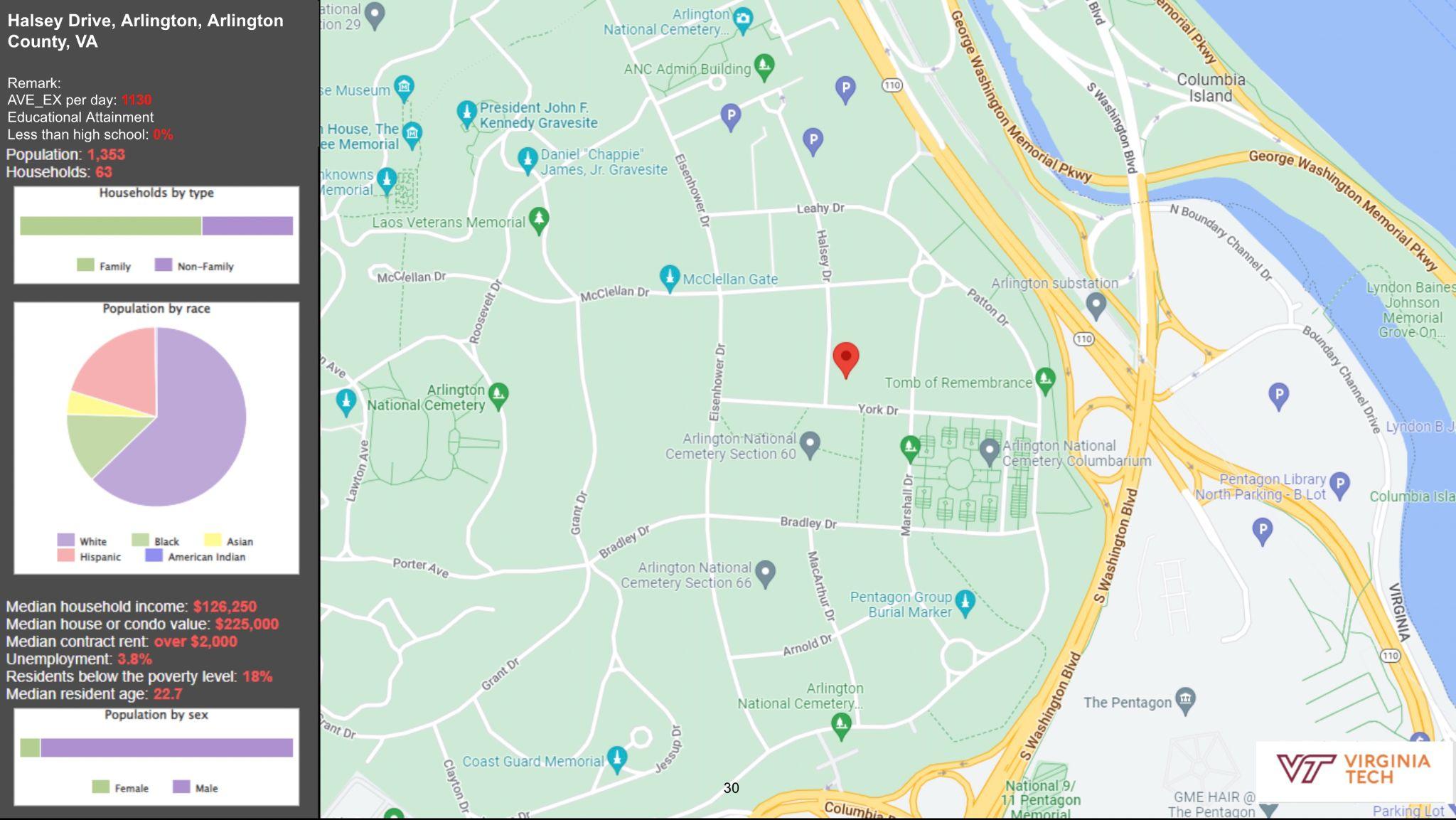
****

Fig 14: Infographic of Arlington national cemetery, VA

A very interesting location emerged in our analysis Arlington national cemetery saw a huge increase around 1K more people visited this place on an average as compared the same period prior to covid. Here in Fig 14 we can observe that the male population in this region outnumbers the female population. We see a mixture of white, black, hispanic and asians in this region, with the average resident age being around 22 years which is quite young.

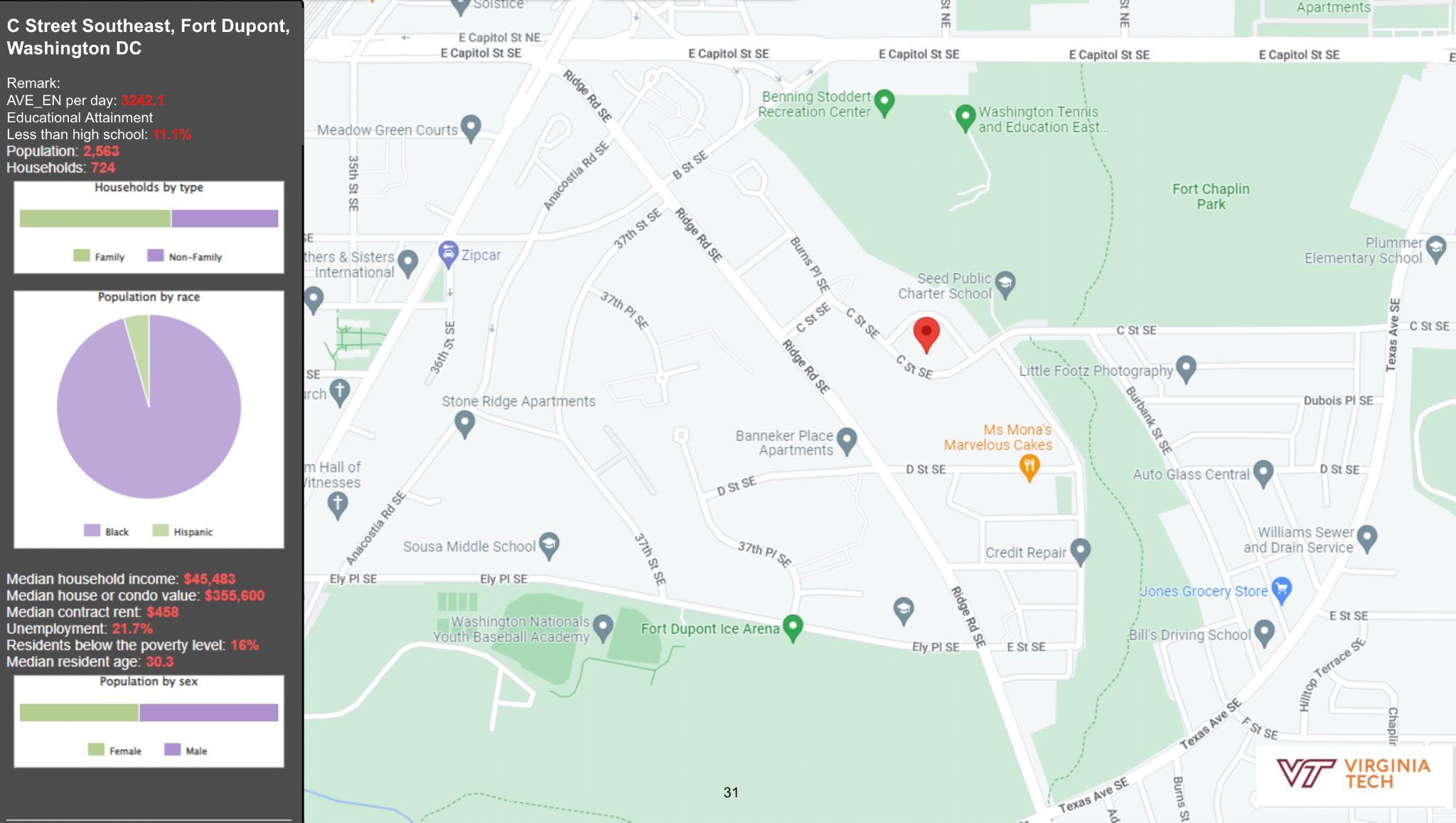
****

Fig 14: Infographic of Fort Dupont, DC

In fig 14 we discuss Fort Dupont in Washington DC. The population is predominantly black around 90%. The number of people not completing high school was around 11%, with average household income of 45,000$ and an unemployment rate of staggering 21%.

Arlington National Cemetery saw a huge increase around 1,000 more people visited this place on an average as compared the same period prior to covid. Trinidad saw the most increase 11,000 more compared to previous year. Brightwood Park, Anacostia DC, Bluemont VA where the percentage of black people was 60-98%, around 8-11% hispanic people saw an average increase of 4,000 people traveling. 22% of the people living in this area have not completed high school, the average household income in this area is 50,000$ and the unemployment rate is 37%

**4.2: Trip length analysis**

| Place | 2020 | | | | |  | 2021 | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 20-Jan | 21-Jan | 22-Jan | 23-Jan | 24-Jan | 25-Jan | 26-Jan | 27-Jan | 28-Jan | 29-Jan |
| UMD College Park, MD | 9.3493 | 9.2074 | 9.0634 | 9.0114 | 9.0169 | 7.0423 | 7.9405 | 7.9744 | 7.6013 | 6.8117 |
| American University Park, MD | 4.5679 | 4.0743 | 3.8090 | 4.0245 | 4.1929 | 4.7817 | 5.1723 | 2.2787 | 4.0373 | 4.1421 |
| George Washington Univ, DC | 5.2920 | 4.5318 | 4.7194 | 3.5712 | 4.4684 | 3.1473 | 3.7767 | 3.7533 | 4.6011 | 3.7340 |
| JF Kennedy Center for Arts, DC | 4.7874 | 4.8871 | 4.7021 | 5.5309 | 4.5849 | 4.7558 | 3.7486 | 3.7035 | 4.0244 | 3.5959 |
| Wesley Heights, DC | 5.6838 | 5.6573 | 5.2158 | 5.0601 | 4.8919 | 4.2319 | 3.5710 | 2.5418 | 3.6386 | 4.0580 |
| Vierra Falls Church, VA | 7.6687 | 9.5461 | 7.6487 | 7.9717 | 7.4819 | 6.6896 | 6.5543 | 3.1125 | 6.5397 | 6.9046 |
| Barnaby Woods, DC | 4.5640 | 4.2244 | 4.7409 | 4.1276 | 4.5171 | 4.8407 | 4.6873 | 3.6152 | 4.3406 | 3.8209 |
| Yorktown Blvd, Arlington, VA | 7.1407 | 6.6492 | 6.8706 | 6.9793 | 7.8844 | 6.2051 | 5.5851 | 4.5954 | 5.0417 | 5.8714 |
| Bluemont, Arlington, VA | 5.3993 | 5.2876 | 5.1788 | 5.7531 | 5.0827 | 4.5697 | 4.5322 | 4.3005 | 4.6731 | 5.1839 |
| Forest Glen, Arlington | 4.4046 | 5.7479 | 5.5590 | 5.7799 | 5.8847 | 4.8832 | 6.2781 | 2.5584 | 4.3146 | 5.6706 |
| Penn Quarter, DC | 4.0525 | 4.5277 | 3.9064 | 4.2716 | 4.0902 | 3.2108 | 2.9362 | 2.8646 | 3.6262 | 4.6585 |
| RR Airport, VA | 9.2728 | 8.3798 | 8.4396 | 9.2315 | 10.683 | 7.6649 | 8.0857 | 2.8157 | 8.1697 | 7.9487 |
| Golden Triangle, DC | 6.3204 | 5.2655 | 5.8059 | 5.5256 | 5.4219 | 5.2952 | 6.4810 | 4.2797 | 5.5668 | 4.5376 |
| The George Univ, DC | 5.6636 | 5.4719 | 5.2543 | 5.2504 | 5.3250 | 3.7453 | 4.3626 | 3.6166 | 4.5199 | 4.6380 |
| Union station, DC | 5.6813 | 5.7221 | 5.7204 | 4.8123 | 5.4258 | 4.1669 | 4.7408 | 2.9772 | 4.9717 | 4.6729 |
| Southwest Federal Center, DC | 4.7406 | 3.7708 | 4.2887 | 3.8428 | 3.6218 | 4.6185 | 6.0832 | 3.9285 | 5.2217 | 5.0161 |
| Downtown, DC | 5.2038 | 5.1004 | 5.1355 | 5.1452 | 5.1084 | 3.5907 | 3.8559 | 4.0065 | 4.3603 | 4.3007 |
| Downtown Silver Spring, MD | 4.8101 | 5.2475 | 4.6363 | 5.0311 | 5.0055 | 5.0393 | 5.1469 | 2.5754 | 5.4053 | 4.4698 |
| Crystal City, VA | 4.9165 | 5.3148 | 5.0926 | 4.7861 | 4.2143 | 4.5063 | 4.9586 | 2.7405 | 3.2404 | 3.5331 |

Table 1. Heatmap of average trip length on weekdays (2020,2021)

We divided the data into weekdays and weekends as we observe the patterns of traveling on weekdays and weekends are completely different as shown in table 2. In table 1 we have a heatmap of the average trip length on the weekdays. We compared the average trip length of the weekend of 2020 (pre-covid) and weekend of 2021 (post-covid). In UMD college Park, MD we observe a massive decrease in the average trip length. On a friday the average trip length was around 9km and after covid its around 6.8km which is a decrease of 2.2 Km. We observe a similar decrease in trip length on monday at Ronald Regan airport and Union station, DC from 9.27 to 7.66 and 5.6 to 4.1 respectively.

| Places | 2020 | |  | 2021 | |
| --- | --- | --- | --- | --- | --- |
| 25-Jan | 26-Jan | 30-Jan | 31-Jan |
| UMD College Park, MD | 10.7014 | 8.7199 | 6.9106 | 6.3824 |
| American University Park, MD | 4.0215 | 4.1961 | 5.2350 | 3.1277 |
| George Washington Univ, DC | 4.8159 | 4.8433 | 3.8591 | 2.9592 |
| JF Kennedy Center for Arts, DC | 4.2703 | 5.3706 | 3.9543 | 3.0753 |
| Wesley Heights, DC | 4.7208 | 5.0774 | 3.8429 | 4.8782 |
| Vierra Falls Church, VA | 8.6443 | 6.7478 | 7.3753 | 3.5599 |
| Barnaby Woods, DC | 5.0931 | 4.5376 | 3.9258 | 4.2122 |
| Yorktown Blvd, Arlington, VA | 5.9929 | 6.6765 | 5.6036 | 3.8336 |
| Bluemont, Arlington, VA | 5.3315 | 5.3521 | 5.1122 | 4.1459 |
| Forest Glen, Arlington | 5.3519 | 5.6531 | 3.4637 | 2.8245 |
| Penn Quarter, DC | 4.2059 | 4.5315 | 4.5121 | 3.1020 |
| RR Airport, VA | 10.1702 | 11.9601 | 7.0129 | 7.3078 |
| Golden Triangle, DC | 5.0114 | 5.1773 | 5.1445 | 4.7720 |
| The George Washington University, DC | 4.8345 | 5.0242 | 5.1110 | 3.1954 |
| Union station, DC | 4.5768 | 5.1884 | 5.3062 | 5.4127 |
| Southwest Federal Center, DC | 3.0574 | 3.4441 | 3.3424 | 4.2739 |
| Downtown, DC | 5.0364 | 5.7011 | 5.1388 | 5.3057 |
| Downtown Silver Spring, MD | 4.5381 | 4.9469 | 4.7790 | 3.0430 |
| Crystal City, VA | 4.8742 | 4.4092 | 4.5123 | 3.5657 |

Table 2. Heatmap of average trip length on weekend (2020,2021)

In table 2 we have a heatmap of the average trip length on the weekend. We compared the average trip length of the weekend of 2020 (pre-covid) and weekend of 2021 (post-covid). In UMD college Park, UMD we observe a massive decrease in the average trip length. On a saturday the average trip length was around 10.7km and after covid it was around 6.9km which is a decrease of 3.6Km. We observe a similar decrease in trip length on Sunday at Ronald Regan airport with average trip length of 11.9 km precovid and 7.3 after covid a reduction of around 4.6km on an average. We also see a few place with slight increase like Union station and downtown.

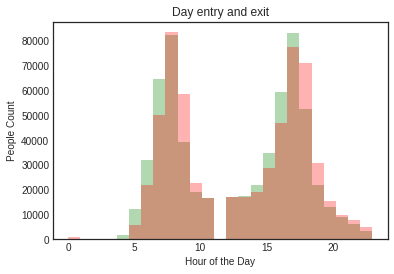
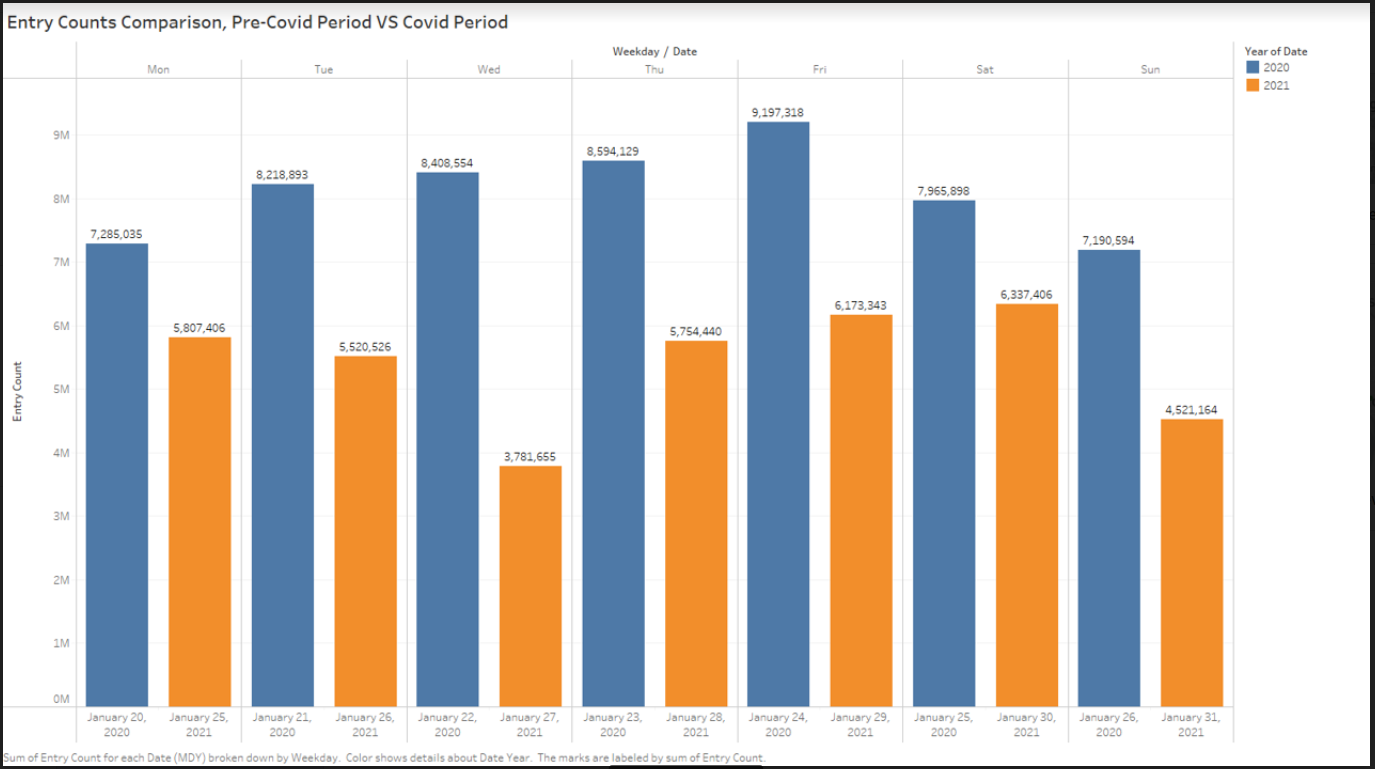
 

Fig 15: Number of entry and exit per hour in smart trip data using bar graph

Total number of entries per day pre covid Vs post covid using bar graph

In fig 15 the green bars represent the entries, the red bars represent exits in each of the stations and the brown bars represent the common region where there is both entry and exit. We observe the number of people taking the bus increase from 6 am and peaks at 10am. We also see at 12 there are very few people using the metro. Similarly we see the increase in people taking the metro back from downtown to subbers increases between 3pm and peaks and 7pm and reduces gradually after 7pm.

We compare total numbers of entries per day pre covid virus total numbers of entries per day post covid in the Fig16. We observe that the total number of people traveling on weekdays was constant pre covid but after covid we can see a dip in the number of people traveling on wednesday. In general there is definitely a decrease in the number of people traveling through the week.

**5. References**:

1. Advani, M.; Sharma, N.; Dhyani, R. Mobility Change in Delhi Due to Covid and Its' Immediate and Long Term Impact on Demand with Intervened Non Motorized Transport Friendly Infrastructural Policies. Transport Policy 2021, 111, 28–37.

2. Alfredo Aloi; Borja Alonso; Juan Benavente; Rubén Cordera; Eneko Echániz; Felipe González; Claudio Ladisa; Raquel Lezama-Romanelli; Álvaro López-Parra; Vittorio Mazzei; et al. Effects of the Covid-19 Lockdown on Urban Mobility: Empirical Evidence from the City of Santander (spain). Sustainability 12 (3870), 3870–3870 DOI: 10.3390/su12093870.

3. Khmaissia F; Haghighi P.S; Jayaprakash A; Wu Z; Papadopoulos S; Lai Y; Nguyen F.T. An Unsupervised Machine Learning Approach to Assess the Zip Code Level Impact of Covid-19 in Nyc. Arxiv 2020, (2020 06 10).

4. Kuo, C.-P.; Fu, J. S. Evaluating the Impact of Mobility on Covid-19 Pandemic with Machine Learning Hybrid Predictions. Science of the Total Environment 2021, 758 DOI: 10.1016/j.scitotenv.2020.144151.

5. Song H.Y; You D. Modeling Urban Mobility with Machine Learning Analysis of Public Taxi Transportation Data. International Journal of Pervasive Computing and Communications 2018, 14 (1), 73–87 DOI: 10.1108/IJPCC-D-18-00009.

6. Qi G; Huang A; Guan W; Fan L. Analysis and Prediction of Regional Mobility Patterns of Bus Travellers Using Smart Card Data and Points of Interest Data. Ieee Transactions on Intelligent Transportation Systems 2019, 20 (4), 1197–1214 DOI: 10.1109/TITS.2018.2840122.

7. Vidovic K; Mandzuka S; Brcic D; 59th International Symposium ELMAR, ELMAR 2017 59 2017 09 18 - 2017 09 20. Estimation of Urban Mobility Using Public Mobile Network. Proceedings Elmar - International Symposium Electronics in Marine 2017, 2017-september, 21–24 DOI: 10.23919/ELMAR.2017.8124426.

8. Aparicio, J. T.; Arsenio, E.; Henriques, R. Understanding the Impacts of the Covid-19 Pandemic on Public Transportation Travel Patterns in the City of Lisbon. Sustainability 2021, 13 (15), 8342–8342 DOI: 10.3390/su13158342.

9. Paiva, S.; Corcoba, V.; Mourao, F.; Paneda, X. G.; Melendi, D.; Garcia, R. Analysis of Mobility Changes Caused by Covid-19 in a Context of Moderate Restrictions Using Data Collected by Mobile Devices. Ieee Access 2022, 10, 8906–8915 DOI: 10.1109/ACCESS.2022.3141083.

10. Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, №34, pp. 226–231).

11. Kang, Y., Gao, S., Liang, Y. Li, M., Rao, J. and Kruse, J. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. Scientific Data 7, 390 (2020). https://www.nature.com/articles/s41597-020-00734-5