DSO 545 Project Report

**Process Overview**

We will next introduce our process of approaching this problem.

First, we will be collecting the data that we require from data sources such as SimplyMap And Census ACS Demographic Data.

Second, we will try and recognize the outlier from the data we collected as same data might have a large amount of public art work, lower income or employment due to certain facilities being in that area.

Third, we will start running regression model with the variable that we currently have.

Forth, we will do a cluster analysis with the data we got from the previous analysis.

Late, we will look into the regression model and start analyzing the results while making hypothesis of why these results are occurring.

**Data Description**

We first research on how to evaluate the economic impact of art pieces and what variable we should look at when evaluating. One research at Princeton, “How the Arts Impact Communities”, suggested a couple of indicators when evaluating the effectiveness of art. We will briefly describe the variables it suggested and use to formulate a model in relation with number of public art pieces.

1. Total Population – The total population within various zip code
2. Total Household Income – The total household income within various zip code
3. Unemployment number and rate – Both the total unemployment rate within various zip code and the percentage of unemployment comparing with total population in that zip code
4. Per Capita Household Income – The household income per capita within various zip code
5. # of Employment in Art & Entertainment – The total number of employee who are in art or entertainment industry
6. Population of high school graduate or higher – The total population who graduate from high school or higher

**Outlier Treatments**

Out of the data we collected we recognized two outliers we will point out and to remove: area 90045 and area 90011.

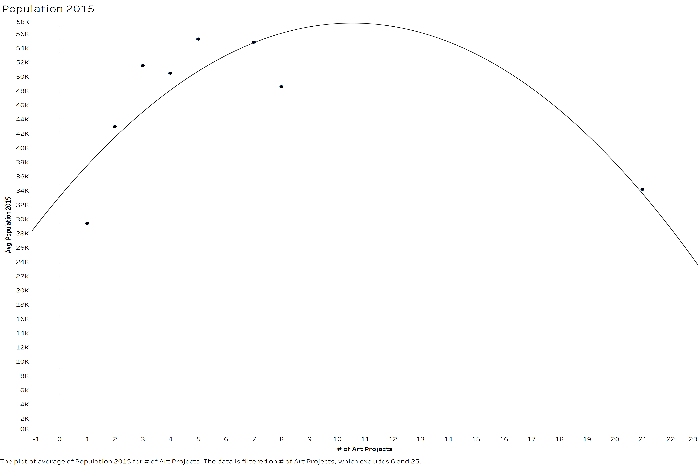
Area 90045: This is where LAX international airport is located and this area has 25 public art. Since other area does not have similar facility (which is an international airport), we will view this as an outlier.

Area 90011: Having USC, a school facilities, we believe the area will have more students who are unemployed and with lower income comparing with other area. Thus, we will exclude this data from our data source.

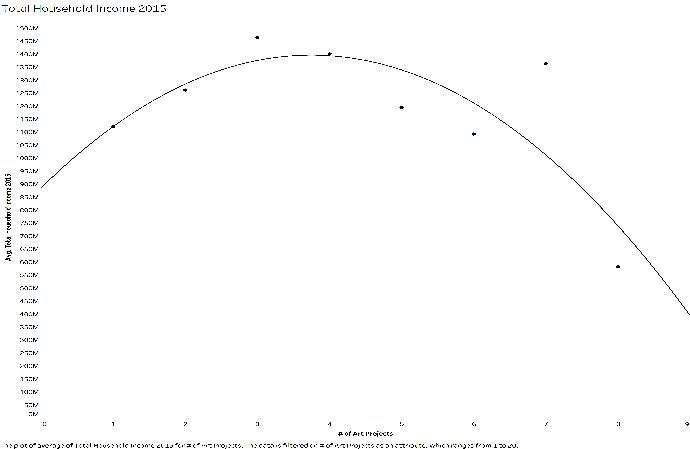
**Data Analysis**

We have data from both 2010 and 2015. By analyzing and compare the results of both we found that they have similar trend line and thus we will be using the latest data of 2015 to develop our regression mode.

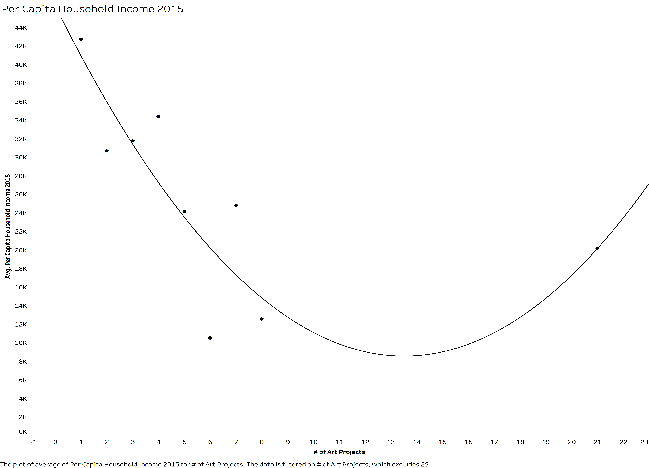
* **Population Analysis of 2015**

For the population analysis, we use a polynomial trend model of a degree of 2 to compute the average population of 2015 given the number of Art Projects. We come up with a significant model at a p-value of 0.048. In the model, we found out that when we have less than 10 art project within an area every increase in art project will have more population. When the art projects exceed 10, every increase in art projects there will be less population.

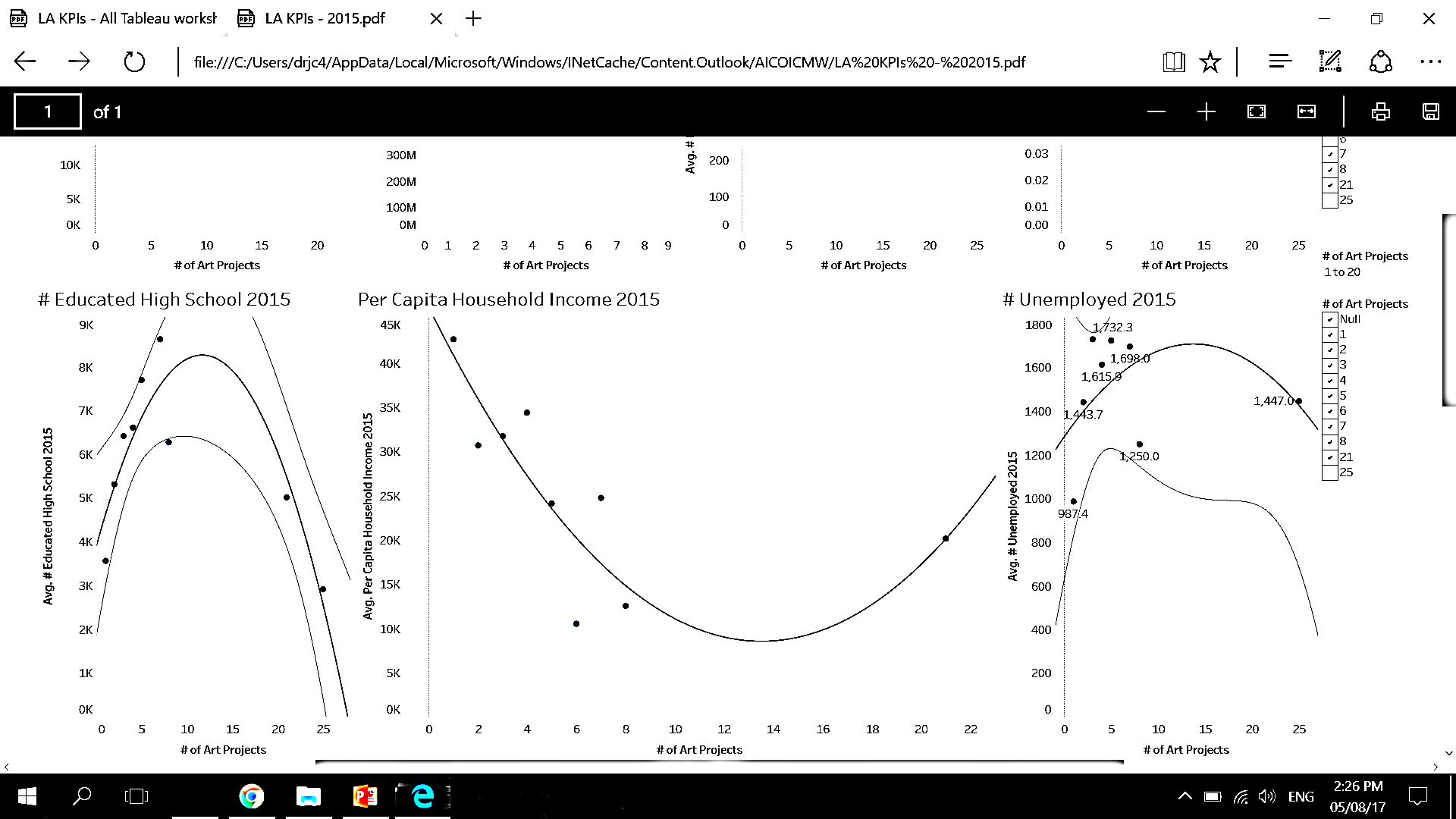
* **Average Total Household Income Analysis of 2015**

For the average total household income analysis, we used a polynomial trend model of a degree of 2 to compute the average of total household income in 2015 given the number of Art Projects. We were able to get a regression model with a p-value of 0.075. In this model, we found that when an area has less than 4 art projects, every increase in art project the average total household income will increase. When the art projects exceed 4, every increase in art projects there will be less average total household income.

* **Per Capita Household Income Analysis of 2015**

For the per capita household income analysis, we used a polynomial trend model of a degree of 2 to compute the per capita household income in 2015 given the number of Art Projects. We come up with a significant model at a p-value of 0.021. In this model, we found that when an area has less than 13 art projects, every increase in art project the per capita household income will decrease. When the art projects exceed 13, every increase in art projects there will be more per capita household income.

* **Number of High School Graduate or Higher Analysis**

For the number of high school graduate analysis, we used a polynomial trend model of a degree of 2 to compute the number of high school graduate or higher in 2015 given the number of Art Projects. We come up with a significant model at a p-value of 0.012. In this model, we found that when an area has less than 11 art projects, every increase in art project the number of high school graduate or higher will increase. When the art projects exceed 11, every increase in art projects there will be less number of high school graduate or higher.

* **Average of % Unemployment Analysis**

A close up of a map

Description generated with high confidenceFor the average % unemployment analysis, we used a linear trend model to compute the average % unemployment in 2015 given the number of Art Projects. We come up with a significant model at a p-value of 0.005. In this model, we found that every increase in art project the average % unemployment will decrease.

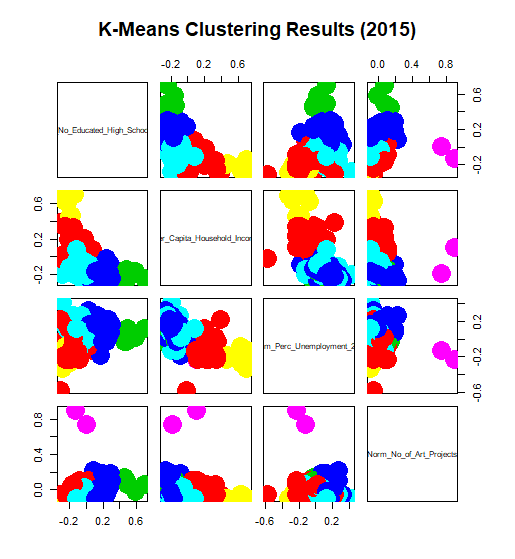
* **Number of Unemployment Analysis**

For the number of unemployment analysis, we used a polynomial trend model of a degree of 2 to compute the number unemployment in 2015 given the number of Art Projects. We come up with a non-significant model at a p-value of 0.67. Thus, we will not focus on the number of unemployment since it is not significantly correlated with number of art projects.

* **Number of Employment in Art & Entertainment Analysis**

For the number of employment in art & entertainment analysis, we used a polynomial trend model of a degree of 2 to compute the number of employment in art & entertainment in 2015 given the number of Art Projects. We come up with a non-significant model at a p-value of 0.14. Thus, we will not focus on the number of employment in art & entertainment since it is not significantly correlated with number of art projects.

**Cluster Analysis**

To further confirm that the number of art projects impacts these variables we would do a cluster analysis base on the variables we end up with. We learn that there are two cluster within the group separated by the number of art project within an area. Once the number of art reaches a certain number that area will become a commercialized area. Any number below will only be divide by the total income.

**Summary Analysis**

We found out that most of our regression model shows an inflection point where at a certain number of art projects the correlation will switch between the variables and number of art works. Out of the five significant models, 3 of them starts with a negative correlation (population, per capita income and education).

The 3 model with the most significant are per capita household income, number of high school graduates or higher and % of unemployment and we will make our hypothesis base on the finding we find in these three variables.

**Hypothesis**

The hypothesis we will make from the result we have is we believe that having more art project within an area does create new jobs thus improving the unemployment rate. Furthermore, by having more art project and commercializing the area, tourism job which does not require professional education will be created. We also believe that suburban areas tend to have fewer art project due to those areas have more wealthier households.

**Next Step**

Finally, we recognized step we can take to make our result more accurate and be further convincing. We found that our current data is skewed to the left (meaning we have a lot of area with art projects less than 8). Having more data on area will more art projects will help prove that the correlation between the variables. We will also hope to look deeper into different case studies to determine more variables to assess the impact of art. Mostly importantly, we hope to obtain the data before and after the installation of the art. Currently we can only state that these is a correlation between number of art project and the variables, but with the before and after data, we are able to better determine the cause and effect of such art project. To further build on that, we would also like to understand how long will an art project have effect on a certain area. By doing so, we believe that our model will be better and even more accurate than before.