To: Abbass Sharif
From: Megan Franklin, Jessica Gelzer, Naman Gupta, Kevin Hisey, Melissa Ng
Date: May 9, 2017
Subject: Art Project Write Up
Link: https://marshallproject.shinyapps.io/marshall_project_dashboard/
_____

Introduction

The Department of Cultural Affairs Public Art Division was looking for an analytics model to measure the progress and impact of the Public Works Improvement Arts Program (PWIAP). PWIAP create arts amenities, facilities and services in connection with all City capital improvement projects. The Program's mission is to provide publicly accessible works of art, arts and cultural facilities, and services for the cultural benefit of the City, its citizens and its visitors.

Since its inception 26 years ago, the Program has collected output and outcome data related to artist profile (ethnicity, gender, place of residence, degrees accomplished, artistic merit and experience), project finances (budget, expenditure, contractual employment), artworks (value of assets, locations/council district, launch date), community/stakeholder meetings, and partnered city departments and facilities.

Our goal was to figure out:
- What, if any, KPIs could be used to measure "progress" and "impact"?
- What models could be developed to track impact?
- Prototype a private dashboard
- Recommend potential data to be aggregated and collected for further analysis

In this write up, we will discuss both our approach to tackling this issue and our findings.

Our Approach

To measure the impact of art projects in LA we took a four stage approach:
- Step One: Hypothesized KPIs
  In this step, we brainstormed potential ways art projects can make an impact on a community and found the relevant data using open sources.
- Step Two: Tested KPIs
  Next, using R, we visualized what types of changes occurred in each zip code and over time.
- Step Three: Created Model
  Then we used a regression model to test the significance of different variables. From there, we built a model to predict future impact of projects.
- Step Four: Designed Dashboard
  Finally using Shiny R, we designed an interactive dashboard for the City of LA to easily visualize the impact of their projects.

Step One: Hypothesized KPIs

After careful discussion, we narrowed down the list of potential KPIs into six different hypotheses:
1. Demographics: Do the demographics impact the response to the projects? (Age and Ethnicity)?
2. Housing Prices: How have art investments impacted housing prices in zip code?
3. Crime Rates: How have art investments impact crime in a zip code?
4. Business Growth: How have the number of businesses changed in an area with art investments?
5. Type of Art Project: Does the type of art project impact the success?
6. Neighborhood Pride: Does the success of a project change if the artist is from the same area?

We will tackle each of these different hypotheses in the following sections.

Step Two: Tested KPIs

(1) Demographics: Do the demographics impact the response to the projects? (Age and Ethnicity)?

First, we wanted to make sure to incorporate demographic data for the project. The data is filtered by zip code and it shows which projects were done in each zip code. The first graph shows the ethnic breakdown for the zip code. The second graph shows the age breakdown for the zip code. The third shows the median income distribution throughout Los Angeles and highlights in red where that zipcode stands. In this example, you can see 90037 is a younger, low-income Hispanic community that had three project investments.

While we did not incorporate it into the model at this point, the demographic information could be useful for future analysis of to meet certain criteria. The information can be used to show equitable distribution of investments across Los Angeles as it relates to ethnicity, age, and income. As the city tries to serve all of its residents, it can make sure it is meeting certain criteria before investing.

(2) Housing Prices: How have art investments impacted housing prices in zip code?

Next, we wanted to incorporate housing prices data and see how they changed over time. We found a dataset for the median home prices of single family residences by Los Angeles County Zip Codes for years 2011-2015 on the Los Angeles Almanac website.[1]

The data is filtered by zip code and highlights that while most housing prices in most zip code increased over time, there were some cases where the line graph did not have a strictly linear relationship.

This dataset was combined with other datasets (crime rate, new businesses, new projects, and project count) to create a model which we will discuss in greater lengths in a later section.

In order to create this new dataset, we filtered out each of the housing prices by year and merged the datasets together.

(3) Crime Rates: How have art investments impact crime in a zip code?

Next, we wanted to look at crime rates over time. This was a little tricky to find because while we were hoping to find a large dataset we also wanted it to be representative of the right areas. We found a dataset from the Los Angeles Open Data Platform which chronicled all of the crimes reported in the City of Los Angeles from 2010 to Present.[2]

The data is also filtered by zip code and highlights how crime rates per zip code have changed over time.

This dataset was combined with other datasets (crime rate, new businesses, new projects, and project count) to create a model which we will discuss in greater lengths in a later section.

In order to create this new dataset, we filtered out each of the housing prices by year and merged the datasets together.

(4) Business Growth: How have the number of businesses changed in an area with art investments?

Next, we wanted to look at new business added over time. We debated over adding the total number of businesses or just the number of new businesses added each year and since we did not have the number of closed businesses per year, we decided to go with the only controlled variable – the number of new businesses added every year.

The data is also filtered by zip code and highlights how crime rates per zip code have changed over time.

This dataset was combined with other datasets (crime rate, new businesses, new projects, and project count) to create a model which we will discuss in greater lengths in a later section.

In order to create this new dataset, we filtered out each of the housing prices by year and merged the datasets together.

(5) Type of Art Project: Does the type of art project impact the success?

Some installations may have more of an impact on others, like freeways and bridges, as they have broader exposure than locations like adult/child care centers. However, many of the locations that are in high traffic locations give the public little time to enjoy the art. Which is more important for art installations, high traffic or time to enjoy the art? For this variable we recoded the variable by hand, using a combination of the project name and the associated organization. For this variable, we created a year over year stacked bar chart, a year to date city wide bar chart, and a chart that filters both year and area.

(6) Neighborhood Pride: Does the success of a project change if the artist is from the same area?

In addition to the project location, we also considered if an artist was local. We felt that this was important to the art installations because local artists may have a local following or connection to a place. In addition, using local artists fosters feelings of pride in the community. In order to create this variable, we web scraped the locations of the zip codes, and then combined these with the existing data sets. From there, we manually recoded the cities into broader locations, based on internet definitions of regions. Finally, we did a match comparing the area of the artist to the area of the project. For this variable we created the same shiny dashboard as location of the installation.

Step Three: Created Model

Using the combined dataset, our team attempted to answer three of the questions listed above:
1. Housing Prices: How have art investments impacted housing prices in zip code?
2. Business Growth: How have the number of businesses changed in an area with art investments?
3. Crime Rates: How have art investments impact crime in a zip code?

To do this, we utilized R to develop 3 regression models with each model having a different Y variable, based on the question we were trying to answer.

What we found was that for each equation – housing prices, business growth, and crime rates – the number of art projects was found to be statistically significant. While this does not absolutely prove that art projects impacts these statistics, it certainly is possible and warrants further examination and research.

Additionally, while each of the models were statistically significant, the model that was examining crime rates was the most significant as it explained nearly 30% of the variance within crime rates. This indicates that the art project program may be best examined through the lens of reducing crime within a given zip code.

Ultimately, however, due to both the immense number of factors that affect the statistics we were trying to explain and the small sample size we were dealing with, it would be foolish and overconfident to definitively conclude and facts from our regression analysis. The key takeaway, however, is that there is a case to be made and more research to be done on number of art projects affecting key stats, namely crime rate.

Step Four: Designed Dashboard

We wanted to create a dashboard which not just outputs the graphs to the user but also is clean enough to be re-usable so Hunter can add more graphs to it later on if needed. In order to build the dashboard, we followed certain steps:

1. We gathered dashboard ideas from open source websites like shiny.rstudio.com and identified the basic elements of various dashboards.
2. We then wrote the R code for each section separately and created graphs for each analysis.
3. The last part was to combine the multiple code bases into one single code file that outputs the final dashboard.

Post-Analysis Conclusion and Next Steps

Recommendations

1. *Foster local talent:*
   When looking for artists for your projects, always consider local talent first. Using local talent fosters pride in the community, and keeps the money invested in the project local, revitalizing businesses.
2. *Make the installation match the location:*
   Art projects should be appropriate to their installation location. For murals on bridges and freeways, people don't have time to enjoy small details. Focus on big, bold, easy to read projects. In contrast, projects in locations like parks need to be more durable, but locals can take the time to appreciate finer details.
3. *Projects have long lasting impacts:*
   As our regression demonstrates, while the number total projects has a statistically significant impact on the number of new businesses and the housing prices. Locations with many art projects are more valuable for home owners and businesses. For both homeowners and new businesses, the number of new projects was not statistically significant, suggesting that the total number of projects is more important than how new they were.

Additional Variables

1. *Business Income*
   If you can obtain local business income, can you measure if the projects have increased foot traffic and thus local business income?
2. *Addresses of Each Location*
   Obtaining addresses of each location could ease the process of mapping the data, and could allow for understanding the radial impact.
3. *Measure Dollar Amount per Impact*
   Using more sophisticated data modeling, it would be worth asking how much money invested creates impact? For every $10,000, do you see an incremental impact and at what point are there diminishing returns?
4. *Type of Installation*
   One additional variable that we would like to measure associated with the projects is the type of installation. For example, is the art project a statue, a mural, or some other kind of installation? Which installations have the strongest impact?

Conclusion

Throughout the project we were able to find some interesting relationship with the variety of information and datasets we had access to. For the demographics hypothesis, while we were able to graph the demographic breakdown per zip code we struggled to find a way to measure the reception of the projects and therefore could not accurately test our first hypothesized KPI. With local pride and type of art installations, too, without an accurate to measure the reception of the project – it was difficult to test the hypothesized KPI.

However, with housing prices, new businesses, and crime rates since project counts are significant in predicting these values – it would be interesting the explore this relationship further by perhaps throwing in more variables to help assess the value add of these various projects.

Sources:
1. "Median Home Prices - Single Family Residences." *Median Home Prices for Single Family Residences by Zip Code in Los Angeles County, California for Los Angeles County, California*. N.p., n.d. Web. 09 May 2017. ‹http://www.laalmanac.com/economy/ec37b.php›.
2. "Crime Data From 2010 to Present | Los Angeles - Open Data Portal." *Data.lacity.org*. N.p., n.d. Web. 09 May 2017. ‹https://data.lacity.org/A-Safe-City/Crime-Data-From-2010-to-Present/y8tr-7khq›.