|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | |  | |  | |  | |  | |
| Net Economic impact: Introduction of Street Car in Downtown Cincinnati  Exploratory Data Analysis, Predictive Analysis and Forecast |
| **K. Rajesh Jagannath**  **Foundations of Data Science** Mentor: Anirban Ghosh 08/05/2016 |

Table of Contents

Net Economic impact: Introduction of Street Car in Downtown Cincinnati 1

Exploratory Data Analysis, Predictive Analysis and Forecast 1

Introduction 1

Objective 1

Motivation 2

Data Sources 3

Data 3

Extraction, Translation and Loading of Data 5

Feature selection 5

# Net Economic impact: Introduction of Street Car in Downtown Cincinnati

## Exploratory Data Analysis, Predictive Analysis and Forecast

Suspendisse potenti.

# Introduction

[The Cincinnati Streetcar](http://www.cincinnati-oh.gov/streetcar/design-route/) is a modern streetcar system designed to link major employment centers in downtown and uptown, connecting through Cincinnati's historic Over-the-Rhine neighborhood.

It will operate 18 hours a day, 365 days a year.

## Objective

The goal of the study is to analyze and predict the “net positive effect” on economy in the buffer zone around the streetcar route by selecting meaningful features from various data sets

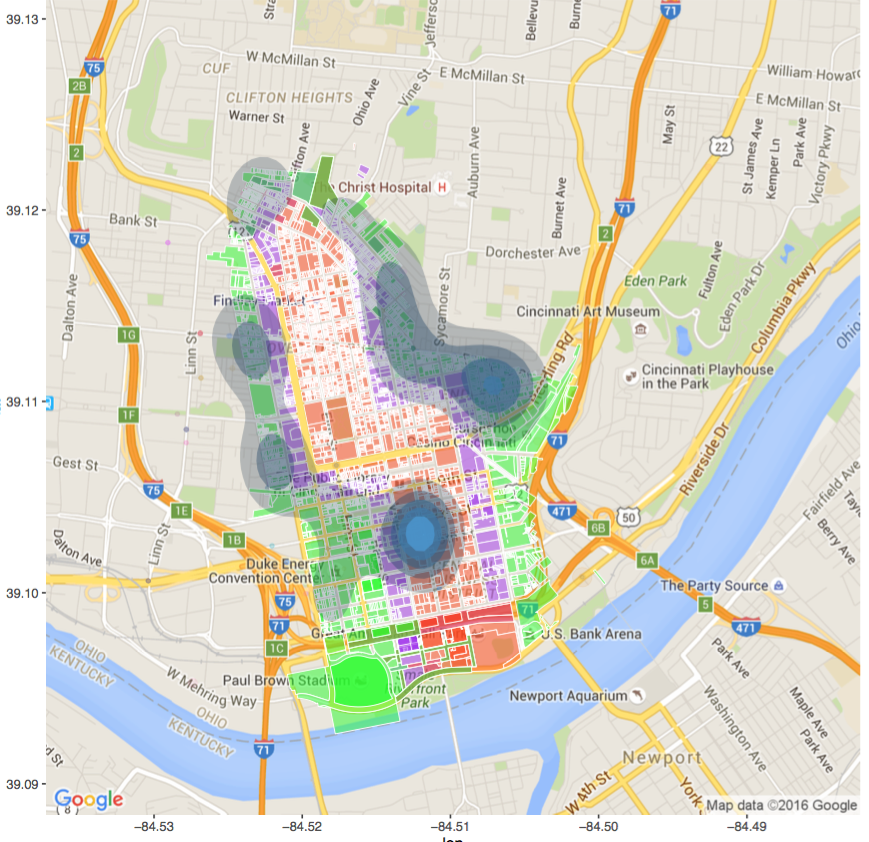
## Motivation

The City of Cincinnati would be the client. Downtown is Cincinnati’s largest employment center, with approximately 70,000 people in the area every day. It has been proven in cities from Atlanta to Seattle that fixed rails in the ground with thousands of potential riders **draw new storefronts and businesses**, as well as housing. That new development will put people to work and **boost the city’s tax revenue.** Also, here may have been inconveniences to the neighborhood, during the construction phase.  Hence, there are two camps of opinion -

* One opinion insists that the introduction of the streetcar is disruptive to the neighborhood ( crowding, transient population, noise) and
* The other opinion is that it provides easy access to business, shops, dining and commuting to work and home and draws new business, expansion of storefronts, revenue from ridership, permit fees, property tax and restaurant license fee.

Three Buffer zones around the streetcar route has been established as shown below. The figure below is the plot of the downtown Cincinnati.

* The area shown in Red color is deemed the CORE Buffer zone. The Streetcar runs through the center of this area on a North South corridor.
* The area shown in Magenta color is designated CENTER Buffer zone and finally
* The area in Green color is EDGE Buffer zone



Data and Sources

## Data Sources

**Data source** would be from [Open Data Cincinnati](https://data.cincinnati-oh.gov/) and Cincinnati Area Geogaphic Information Systems ([CAGIS), City of Cincinnati, OH.](http://cagismaps.hamilton-co.org/cagisportal)

1. **Study Area Parcels : There are three .csv files** . The PARCELID will uniquely identify data in other files if the parcel is classified as being in CORE BUFFER zone, CENTER BUFFER zone, EDGE BUFFER zone or none at all. The following table features are of interest in the 3 files : Each file has about 1000-2000 observations

|  |  |  |  |
| --- | --- | --- | --- |
| **Column name** | **Example Data** | **Description** | **Source** |
| **PARCELID** | 7500010007 | Unique id to identify parcels | StreetCarParcels\_CORE.csv  StreetCarParcels\_CENTER.csv  StreetCarParcels.EDGE.csv |
| **EXLUCODE** | C | Existing Land use Code |
| **MKTLND** | 28180 | Market Value of the Land |
| **MKTIMP** | 90810 | Market value of the Improvementts |
| **ADDRNO** | 1208 | Address, street and type of street |
| **ADDRST** | SYCAMORE |
| **ADDRSF** | ST |

**Table 1: .csv files are used to identify the parcel id. of the three areas around the Street Car - Core, Center and Edge Buffer zones**

1. **Tax Information 2007-2015 .fwf :** Data was provided for 9 years in Fixed Width Format in 9 files. This Dataset needs to be converted to .csv format. The key feature that will be subset from this is parcel\_id and Tax information. A time series analysis on the taxes will be performed to forecast the revenues to the City of Cincinnati. Each file has 30,000 observations and is at leat 300 MB is size

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Example Data** | **Description** | **Source** |
| **SMDA\_NUM** | 10001000100 |  | Data2/taxinfo/2007/taxinfo2007.csv  .  .  .Data2/taxinfo/2015/taxinfo2015.csv |
| **UNIT** |  |  |
| **LOC\_STREET** |  | Location e.g for 2327 Sussex Ave., Cincinnati Oh |
| **LOC\_HOUSE\_NO** | 2327 |
| **LOC\_ST\_DESC** | SUSSEX |
| **LOC\_ST\_IND** | AV |
| **LOC\_ST\_DIR** |  |
| **VALID\_SALE** | Y | Yes or No |
| **MKT\_LAND\_VAL** | 23000 | Value of the Land |
| **NUM\_PARCEL** | 3 | Number of Parcels |
| **DEED\_TYPE** |  |  |
| **MKT\_IMPR\_VAL** | 140570 | Market value of the Land |
| **SALE\_AMOUNT** | 116000 |  |
| **SALE\_DATE** | 20121129 |  |
| **MKT\_TOTAL** | 163570 | Mkt Total Val |
| **NEW CONSTR** | N | Was this a newly constructed building |
| **ANNUAL\_TAXES** | 3693.14 | Annual Taxes Billed |
| **FRONT\_FOOTAGE** | 7000 | Front footage |
| **ACRES** | 0.246 | Acreage of the building |
| **TAXES\_PAID** |  | Annual Taxes Paid |
| **PARCEL\_ID** | 10001000100 | Unique identification number of parcels |
| **DELQ\_TAXES** | 6088.56 | Delinquent taxes |

**Table 2. Features selected from Property Tax Information from years 2007 - 2015**

# Extraction, Translation and Loading of Data

**For now we will be looking at the features from files mentioned in Tables 1 and 2.**

1. **Study area** : Data around the Street car route is divided into 3 Buffer zones. The StreetCarParcels\_CORE.csv, StreetCarparcels\_CENTER.csv and StreetCarparcels.csv files identify the parcels belonging to the zone. The unique identifier is parcel id. In the taxinfo files, the PARCELID will be used to classify/ subset an observation to be belonging to CORE, EDGE or CENTER areas. For the study area preliminary exploration of the data was done : https://github.com/krajeshj/Streetcar0719/blob/master/ExploratoryDataAnalysis.md

2. **Property Tax information 2007-2015** : The original datasets were provided in fixed width format. An R script converted it to .csv file. The problem here was each of the groups of years 2007, 2008 and 2009-2014 and 2015 had different column widths. Luckily, there was clear documentation of field widths. As a result, the files have been successfully converted to .csv file.

# Feature selection

The features that will be selected for analysis will be:

PARCELID – *A Unique identifier of the parcel in the Buffer zone*

+

ADDRNO

ADDRST *- Address for plotting on ggplot or other package to identify spatial correlation*

ADDRSF

+

EXLU\_CODE *– Existing Land*

*PLUS*

for each\_year in (c(seq(2007:2015)))

MKT\_LAND\_VAL

MKT\_IMPR\_VAL *– Market Value of land, Improvements, and Total*

MKT\_TOTAL

+

ANNUAL\_TAXES

TAXES\_PAID - *Net Prop Tax revenue: indicator of growth in economy*

DELQ\_TAXES

+

FRONT\_FOOTAGE

ACRE - *Sq. Footage or acreage to compute Property val/sq. foot*

+

SALE\_AMT

VALID\_SALE

SALE\_DATE - *Stochastic data on valid sale date*

NEW\_CONSTR

+

DEED\_TYPE - *Type of Deed for Classification*

+

DUMMY\_VARIABLE \_CENTER

DUMMY\_VARIABLE\_EDGE

DUMMY\_VARIABLE\_CORE *– Dummy variables of Classification*

DUMMY\_VARIABLE\_NONE

end

Exploratory Data Analysis :

* **Exploration of Project Data : [https://github.com/krajeshj/Streetcar0719/blob/master/ExploratoryDataAnalysis.md](https://github.com/krajeshj/Streetcar0719/blob/master/ExploratoryDataAnalysis.md" \t "_blank)**

Build Models for Time Series analysis

Visualization of Forecast data

Code on github

Bibliography

References

1. <https://dev.socrata.com/foundry/data.cincinnati-oh.gov/emnx-rw6d>
2. <http://www.cincinnati.com/story/news/2016/05/05/streetcar-nation-kc-opens-friday-cincy-next/83874740/>