

Data Alliance for Science
Dept. of Data Science, University of Washington
deepa15@uw.edu
orbite@uw.edu
anqiw2@uw.edu

Seattle Gentrification Atlas

Prepared by:

Deepa Agrawal
Erin Orbits
Angel Wang

Prepared for:

Bernease Herman, Data Scientist, eScience Institute
Rachel Berney, Assistant Professor in Urban Design and Planning, College of Built
Environments
Gundula Proksch, Associate Professor, Architecture Department

Table of contents

Table of contents	2
Statement of Problem	3
Design Objectives	3
Project Management	4
Communication and Coordination with Sponsor	4
Team Qualifications	4
References	5
Appendix A:	5
Résumés of Team Members	5

Statement of Problem

Seattle is one of the fastest growing big cities in the U.S. While prosperity provides economic opportunities to many, it also exacerbates urban inequity. Gentrification is one major process that contributes to inequality. From different disciplines, a gentrifying neighborhood is one that moves from the bottom half to the top half in the distribution of median household income, rental prices, and/or home prices.

This project will produce a Gentrification Atlas, which allows stakeholders in the city's development to identify the neighborhoods in danger of gentrification and visualize trends and impact of potential changes in Seattle. The visualization tool can also be of use to policymakers, researchers, and citizens.

The primary task for the project is to create a prediction model of gentrification, utilizing primarily Census data from the past two decades such as the median income of households in each neighborhood. With the gentrification definitions from social sciences, financial and economics domain, the project is also intent to identify which neighborhoods have gentrified and which are currently gentrifying and which neighborhoods that are most in danger of gentrifying in the near future from different perspectives.

In some respects, this project is a continuation of the earlier Equitable Futures project administered by the University of Washington eScience Institute during the annual Data Science for Social Good (DSSG) program. The Equitable Futures project developed:

- (1) a prototype for an interactive visualization tool for mapping equity indicators related to housing, income, mobility, and education on the city and neighborhood scale and
- (2) a structural equation model that attempted to measure relationships between publicly available data and underlying socioeconomic factors affecting equity.

The data used in the Equitable Futures project was primarily sourced from the American Community Survey (ACS) and the City of Seattle open data portal. That data, along with the earlier work is available as a resource for the current project.

Design Objectives

1. Conduct exploratory analysis using visualization and unsupervised learning to better understand available data.
2. Operationalize three select definitions of gentrification for use as ground truth in a supervised learning model.
3. Interpret those operationalized definitions, choosing the best data proxies to represent relevant phenomena as features.
4. Train a model that will allow you to analyze and examine the relationships and weights between features.
5. Combine data analysis with GIS-based visualizations to produce spatial maps of gentrification forces in the city.

Project Management

Communication and Coordination with Sponsor

For major milestone meetings with the sponsor, all members will attend the meetings in person (if possible). We will all contribute questions, feedback, and status reports for contact with the sponsor. But, for smaller questions or issues Erin Orbits will be the point contact via phone or email. Depending on the project and the wishes of the sponsor, we may share a GitHub repository or Google document to keep the sponsor apprised of our work.

Following is our communication plan for next 5 months:

- Weekly in person meetings on Thursdays with members of the sponsor group
- Using Slack channel for intra-team status updates and questions
- Online meetings on weekend as needed
- Using a GitHub repository for code and documentation updates

Team Qualifications

Deepa

Deepa is pursuing Masters in Data Science at University of Washington starting Sept' 16. 10+ years of Software Development Life Cycle (SDLC) experience in Requirement gathering, Design, Development, Data Modelling and Business/Data analysis. Multi domain experience in Ecommerce, Insurance, Mobile, Online gaming and Avionics. Planned responsibilities for the capstone project include data acquisition, data processing, maintaining data, query execution and contributing to machine learning model.

Erin

Erin is pursuing a Masters of Science in Data Science at the University of Washington and expects to graduate in March 2018. Prior to enrolling in this program, she was a full-time attorney. Erin is looking forward to applying her statistical analysis, machine learning, and data visualization skills to a wide range of problems. But, she is especially interested in outlier detection problems.

Angel

Angel is an aspiring data scientist with four-year proficiency in data analysis, statistical modelling, project management, and translating statistical results into business recommendations. Industry experience in R and Python for writing machine-learning algorithms, SQL for database querying, Hive/Hadoop for distributed computing, and Tableau for generating visualizations. Angel will be the one who is responsible to leverage machine learning algorithms to build supervised learning model and utilize her background in Statistics to better understand the data.

References

- [1] Bernease Herman et al., Data science for urban equity: Making gentrification an accessible topic for data scientists, policymakers, and the community,
<https://export.arxiv.org/pdf/1710.02447>
- [2] Lance Freeman, Displacement or Succession? *Residential Mobility in Gentrifying Neighborhoods* (March 1, 2005),
<http://journals.sagepub.com/doi/abs/10.1177/1078087404273341>
- [3] Sample Census Infographics & Visualizations,
<https://www.census.gov/library/visualizations.html>

Appendix A:

Résumés of Team Members

DEEPA AGRAWAL

Redmond, WA 98053 | C: 425-449-3295 | deepa15@uw.edu | <https://www.linkedin.com/in/deepa-agrawal/>

PROFILE

- Master of Science, Data Science at University of Washington starting Sept'16.
- 10+ years of SW experience: Requirements, Design, Development, Data Modelling, Business/Data analysis.
- Multi domain experience: Ecommerce, Insurance, Mobile, Online gaming and Avionics.
- Lead experience: led a team of 6 resources in a data warehousing project.
- International work experience: US, Canada and India.

EDUCATION /DESIGNATIONS

Master of Science: Data Science (Current GPA: 3.87) **University Washington, Seattle** **Class of 2018**
Engineering (BTech - IT) **National Institute of Technology, India** **June 2005**

AINS : Associate in General Insurance **The Institutes, United States** **May 2014**

SKILLS

Database: SQLite, SQL Server, SQL Azure **Tools:** SSMS, SSRS, MySQL Workbench, Apache Drill, JDBC

Data Visualization tools: Tableau, Power BI **Languages:** SQL, R, C#, Python

PROFESSIONAL EXPERIENCE (2005-2017)

Microsoft Intern in Azure Networking – Redmond **Jun'17-Sept'17**

As a data science intern, responsible for analyzing Ping Mesh data to improve networking coverage and real-time alerting.

Implementing statistical models to programmatically detect spikes and trends in latency across various dimensions.

Working on fine tuning the existing Stochastic Gradient Descent algorithm to get better confidence level.

Microsoft via Wipro Limited – Redmond **Mar'15-Apr'16**

As a senior engineer, responsible for data analysis of test failures and overall health of the triage pipeline.

Created live dashboards using Power BI to monitor Windows test runs.

Berkley North Pacific (BNP) – Bellevue **Aug'13-Dec'14**

As a Quality Analyst, owned multiple Lines of Business: General Liability, Commercial Auto, Commercial Property, Business Owners Policy and Umbrella coverage.

Used data analysis techniques to validate business rules and identify issues in policy pricing.

Cognizant Technology Solutions – Pune **Mar'10-Aug'10**

A strategic three-way reconciliation for the JPMC to match Balance Sheet Mark to Market, Realized/Unrealized Profit and Loss between numerous Risk systems, Sub Ledgers and General Ledger systems. As a Test Lead, managed a team of 6 resources to complete end to end ETL testing.

Wipro Technologies – Bangalore, Pune, Toronto **Sept'05-Jan'10**

HiPlus for Aviva Hibernian, UK: An Insurance website used by Hibernian Direct Ltd (HDL) for motor, home and travel insurance products.

Smoke Detection Function for Airbus (A-340), Germany: As a software engineer owned testing and verification of the SDF (Smoke Detection Function) Module.

Erin Orbits

Attorney and Data Scientist

orbite@uw.edu | cell (425) 444-6269 | github.com/orbitse

Objective: To utilize and expand my statistical analysis, programming, and data visualization skills working on a Data Science Team where I can also apply my legal knowledge.

TECHNICAL SKILLS

Statistics

Bayesian inference
Descriptive Statistics
Hypothesis Testing

Machine Learning Tools

SciKit Learn
TensorFlow

Machine Learning Algorithms

Gradient Descent
K-means
K nearest neighbors
Logistic Regression
Primary Component Analysis
Ridge Regression

Data Management

AWS: EC2, EMR, Redshift
Azure
Hadoop, Hive
SQL, SQL Server

Programming Languages

Python, R, Java

EDUCATION

University of Washington (Sept. 2016 – Present)

Masters of Science in Data Science, exp. Mar. 2018

Current cumulative GPA 3.7

Practiced applied statistics and experimental design, machine learning techniques, MapReduce, parallel processing, and user-centered data visualization in ggplot2, Matplotlib, Bokeh, d3.js and Tableau

Seattle University Law School (July 2004 – May 2007)

Juris Doctor, *magna cum laude*

Awarded President's Scholarship; CALI Award for highest grade in Appellate Advocacy; and competed on the Moot Court Team, winning Best Advocate Award in national law competition

Whitman College (Sept. 2000 – May 2003)

Bachelor of Arts, with distinction, Economics

Awarded President's Scholarship; Elected Student Body President

RECENT WORK EXPERIENCE

Attorney, Private Practice (Nov. 2012 – Sept. 2016)

Specialized in statutory construction and interpretation; conducted research, drafted briefs, and consulted

Deputy Prosecuting Attorney, Pierce Co. (Feb. 2008 – Oct. 2012)

Tried 40+ cases, including a 3rd Strike, 1st Degree Assault jury trial that resulted in a guilty verdict; Argued appellate cases including the appeal of a juvenile court order dismissing a deferred disposition before the Court of Appeals: State v. D.P.G., 169 Wn. App. 396, 280 P.3d 1139 (2012); Exercised judgment in thousands of negotiations

SELECTED DATA SCIENCE PROJECTS

Data Visualization Tool

Cleaned, standardized, and analyzed WA Dept. of Revenue data on open and closed businesses in Seattle, then merged that data with geolocation data and GIS shapefiles to create an interactive tool for analyzing small business trends in Tableau.

Image Classifier Model

Used TensorFlow to build a convolutional neural network, extracted features from 10,000 photos of birds, and used a logistic regression model for identifying 200 species of birds in an GPU EC2 instance before saving the predictions in a CSV file.

Anqi Wang | 588 Bell St Unit 2705S, Seattle, WA 98121 | anqiw2@uw.edu | (206) 790-5569

PROFESSIONAL SUMMARY

Aspiring Data Scientist with four-year proficiency in data analysis, statistical modelling, project management, and translating statistical results into business recommendations. Industry experience in R and Python for writing machine-learning algorithms, SQL for database querying, Hive/Hadoop for distributed computing, and Tableau for generating visualizations.

WORK EXPERIENCE

Expedia, Inc., Bellevue, WA Jun 2017 – Sep 2017
Data Science Intern

- Loaded and modeled Expedia booking data in Hadoop, and processed the data for analysis to answer customer service experience questions.
- Developed, implemented and tested statistical models using R (caret, party, tree, mboost, e1071, rpart, etc.) and Python (SciKit-Learn, TensorFlow, Scrapy, etc.) to show how booking variables impact customer service contact behavior and recommended business solutions to deflect customer service contact and provided effortless customer experience.
- Project: Created predictive customer service call models utilizing 23M Expedia 2016 booking transaction data. Three predictive models were built and tested in R and Python: identifying likely callers, predicting caller window and caller need. Proposed initiatives on service call reduction and customer segmentation, using personalized experience based on insights from analysis. Customer service calls have been reduced by 4% after implementing the model for test and learn in September 2017.

UW Medicine, Seattle, WA Sep 2016 – Jun 2017
Analyst Intern

- Conducted statistical analysis on employee on-boarding survey data to generate predictive analysis trends using Excel and R. Produced and presented reports, using PowerPoint and Word, to senior leadership in HR to improve employee satisfaction.
- Provided administrative support and managed record maintenance based on established retention schedules for HR.

Wells Fargo, Seattle, WA Jan 2015 – Jun 2016
Financial Advisor/Licensed Private Banker

- Managed client relationships by structuring portfolios catering to clients' financial status, objective, risk tolerance, tax exposure and investment goals.
- Leveraged computational analytics to devise accurate projections of portfolio returns for clients.

Expeditors International of Washington, Inc., Seattle, WA Jan 2014 – Dec 2014
Business Analyst

- Analyzed transportation and insurance trends in data sets using SQL and Excel to improve annual operational efficiency by 20%.
- Proposed solutions aligned with customer needs and increased satisfaction by 7% in Q4 2014.

EDUCATION

University of Washington, Seattle, WA Expected: Mar 2018
M.S. Data Science Current GPA: 3.91

Relevant coursework: Data Visualization & Experimental Design, Data Management for Data Science, Statistical Machine Learning, Scalable Data Systems & Algorithms.

Project: Built a visualization application (TravelViz) to assist users in making informed travelling decisions based on area of interest, neighborhood, time of year and time of day. The application data, over 8GB, was sourced from 2.7M Yelp reviews. Using R, I applied statistical modeling and ANOVA to analyze trends between user reviews and business star rating, and Natural Language Processing to extract keywords from reviews. Visualizations from the analyzed data were generated using Tableau.

Link: <http://cp6863.axshare.com/#c=2>.

University of Washington, Seattle, WA Dec 2013
B.S. Statistics & Economics, Minor in Mathematics GPA: 3.55

QUALIFICATIONS

- Programming Languages: R, Python, SQL, JavaScript, HTML, CSS
- Cloud computing platform: AWS (Redshift, EC2, S3), Azure
- Computing Framework: Hive/Hadoop, Spark
- Data Visualization Tool: Tableau