Project WASP: Geo Spatial-Temporal analysis of construction trends using building permit info.

Problem Statement: There are no good tools to visualize trends over time and analyze how the construction densities have changed in a city (ex.Chicago).

Heilmeier Questions:

1. **What are you trying to do? Articulate your objectives using absolutely no jargon.**

*Our aim is to build an interactive analytics portal to assess construction density trends in an urban area.*

1. **How is it done today; what are the limits of current practice?**

*Currently,*

* 1. *We see static visualizations and don't find trend charts and geo-spatial analysis done together*
  2. *Lot of focus is on residential and/or commercial real estate pricing, however our focus is to understand all types of building permits*

1. **What's new in your approach? Why will it be successful?**

*Our tool would offer,*

* 1. *Dynamic and Interactive Geo-temporal analysis of construction in a city*
  2. *Out of the box visualizations for unsophisticated users and also ability to create a custom view*

1. **Who cares?**

*Our users would be :*

* 1. *Real Estate Analysts - pricing depends on factors like train stations, shopping centers, schools, and hospitals also in the area rather than just new homes.*
  2. *City Planners - Understanding the construction density trend, city planners can make much more informed decision on where and by when to build more public infrastructure to support the population*
  3. *Real Estate ( Buyers/Sellers ) - Get a holistic view of upcoming development in an area to make a more informed decision*

1. **If you're successful, what difference and impact will it make, and how do you measure them (e.g., via user studies, experiments, ground truth data, etc.)?**

*Geo-Spatial visualization will help users gain insights into construction and real estate trends. This tool will help city/authorities make data-driven decisions for infrastructure, sustainable living planning. For, construction companies, this can serve as a great way to visualize how trends have changed over time.  
The tool adoption can be assessed by measuring the traffic to the web page. This can be automated with tools like Google Analytics. For specific use cases for enterprises, user surveys can be conducted to prove the hypothesis/assumptions about the tool.*

1. **What are the risks and payoffs?**  
   ***Risks:***
   * *Data cleaning can get complex and time consuming.*
   * *Geo-Spatial visualization using 3rd party libraries can become risky if they don’t support what we are intending to build.*
   * *Scope of the project might get huge given the amount of time we spend on homework.*

***Payoffs:***

* + *The ability for retail users to make sound financial decisions.*
  + *City/Town planning can use this tool for data driven decisions vs anecdotal instincts.*

1. **How much will it cost?**
   1. Resource Cost (10 hrs/person/week) = 50 hrs/wk \* 8 weeks = 400 resource hours
   2. Storage Cost - Tier 2 Data Lake = 1$/month
   3. Computation Cost – 24.7$/Month (using AWS ec2 pricing of 0.0139/hr)
   4. Miscellaneous – 1% of the overall cost.
2. **How long will it take?**  
   *7 weeks for project to be completed, 1 week for presentation prep.*
3. **What are the midterm and final "exams" to check for success? How will progress be measured?**  
     
   Following is the plan of activities and also serves as checkpoints to track project:

|  |  |  |
| --- | --- | --- |
| Activity | Assigned | Date |
| Data Gathering, Cleanup and Storage | Brian Tran, Aayush Parwal | 03/26/2021 |
| POC for visualization with fixed data | Prashant Kubsad, Wael Sulthan,Scott Bader | 04/09/2021 |
| Time series analysis of data | All | 04/15/2021 |
| Final working tool with complete data and analysis. | All | 04/25/2021 |
| Final Presentation and wrap up | All | 05/01/2021 |

Literature Survey:

Building permits are great information to understand construction trends. A spatio-temporal analysis done in the article[PK-1] provides us, with lots of parallels in our aim to visualize the construction trends over time. Analysis of type of work, value would enrich understanding beyond time and space linked with world events [PK-1]. Lot of work has been done on usage choropleth maps to visualize geo spatial models, like dynamic increase in percievable area[PK-2], boundary neighbour selection [PK-2]. This coupled with Google Maps/API, gives us ability to develop interactive webpages. Reactive time component to geo-spatial models, presents its own challenges. Possible solutions are discussed in EST[PK-3]. We can combine the principles mentioned in EST[PK-3] with web development technologies[PK-4] to provide an easily accessible tool that visualizes trends in construction patterns.

The journal is presenting different ML models like ARIMA and exponential smoothing to enhance transportation system. The paper helped us identify different models and how to use them in predictions. Exploring existing machine learning platform like Google cloud and MS Azure could save us hassle and time[WS-1]. The second paper provides home-seeker an interactive visual system. The variety of visualization included provided us a complete view on different designs and the way of use. The potential challenge is the level of complexity of some graphs yet utilizing third party tools like Tableau and Power BI will solve it [WS-2]. The last paper presents the use of ML algorithms to predict city expansion. The paper integrated two models, the Markov chain and the Cellular Automata. This will help us in forecasting urbanization growth. The availability of satellite images is a challenge. However, we could rely on other relevant public data [WS-3].

Main idea of the study[AP-1] is to address how advances in geo-spatial analysis is influenced by social sciences as construction permits for residential, commercial, or public buildings go hand in hand with socio-economic demography of an area. Given the time, study cites major challenges in being able to read, manipulate and store large amounts of detailed data, which is required for any geo-spatial analysis, like maps, roads etc. Our project aims at using today's accessibility of such data from local governments and combining it with flexibility of cloud making such limitations go away.  
Main Idea of the author here[AP-2] is to identify extent of damage and recovery efforts based on building permits and a spatial scan. This is directly relevant to us as city planners, specifically in danger areas, can use our website/tool to balance giving out building permits to residential, and also understand disaster recovery clusters and allocate resources accordingly. Author touches but fails to fully integrate re-population, or not immigration pre and post disaster. For example, if people chose to move away from the region permanently.  
The author of the study[AP-3] looks to utilize density of population to dynamically adjust k value in the algorithm. This is particularly useful for us as even within city concentration of building permits needs to be changed for example, city center vs suburbs etc. Author fails to show how the cluster moves/trends as external factors influence density. Our analysis using the building permits will look to perform trend analysis on these clusters

Previous research using construction data as basis for analysis, show that such data can yield meaningful insights, in terms of trends and linkages to events [SB-1][SB-5].

We seek to build upon earlier efforts. Some earlier efforts used outdated technology (e.g., ESRI ArcGIS) and outdated methods (e.g., Microsoft Office Excel and Access) to organize data [SB-4]. Other efforts used effective data analytics techniques, but deficient visualizations [SB-2]. We can improve visualization by replacing static diagrams with interactive, dynamic and better practices (avoid red-green color schemes. [SB-2].

We consider various techniques for analyzing [SB-2] [SB-3] [SB-6] our baseline building construction data set.

We consider joining our base data set with complementary data, based on news articles, which we could harvest either by web-scraping or directly using API. Either way, candidate data set would be a corpus of unstructured data. Plan would be to use techniques such as TD-IDF [SB-7], or perhaps more cutting edge methods [SB-8] to analyze the data.

1. The Future of Spatial Analysis in the Social Sciences

Main Idea : The study mainly wants to address how advances in geo-spatial analysis is influenced by social sciences and vice versa.

This is useful as construction permits for residential, commercial or public buildings go hand in hand with socio-economic demography of an area.

Given the time, study cites major challenges in being able to read, manipulate and store large amounts of detailed data which is required for any geo-spatial analysis, like maps, roads etc. With today's accessibility of such data from local governments and combining it with flexibility of cloud, makes such limitations go away around gis.

2. Using Building Permits to Monitor Disaster Recovery: A Spatio-Temporal Case Study of Coastal Mississippi Following Hurricane Katrina

Main Idea : of the author here is to identify extent of damage and recovery efforts based on building permits and a spatial scan.

This is directly relevant to us as city planners, specifically in danger areas, can use our website/tool to balance giving out building permits to residential, and also understand disaster recovery clusters and allocate resources accordingly.

Author touches but fails to fully integrate re-population, or not immigration pre and post disaster. For example if people chose to move away from the region permanently.

3. Adaptive clustering algorithm based on kNN and density

Main idea: The author looks to utilize density of population to sample t dynamically adjust k value in the algorithm.

This is particularly useful for us as even within city concentration of building permits needs to be changed for example, city center vs suburbs etc.

Author fails to show how the cluster moves/trends as external factors influence density as against to a cluster behaving in a silo. Our analysis using the building permits will look to perform trend analysis on these clusters

References:

PK-1 : Using Building Permits to Monitor Disaster Recovery: A Spatio-Temporal Case Study of Coastal Mississippi Following Hurricane Katrina - <https://www.tandfonline.com/doi/abs/10.1559/152304010790588052>

PK-2: Dynamic Choropleth Maps – Using Amalgamation to Increase Area Perceivability

<https://ieeexplore.ieee.org/abstract/document/8564174>

PK-3: Exploratory spatio-temporal visualization: an analytical review

Journal of Visual Languages & Computing, Volume 14, Issue 6, December 2003, Pages 503-541

<https://www.sciencedirect.com/science/article/pii/S1045926X03000466>

PK-4: Data Visualization with D3JS and Angular – Christoph Korner

<https://www.google.com/books/edition/Data_Visualization_with_D3_and_AngularJS/FQOzCAAAQBAJ?hl=en&gbpv=1&dq=d3js+with+angular+application&pg=PP1&printsec=frontcover>

WS-1: Smart transportation planning: Data, models, and algorithms

<https://www.sciencedirect.com/science/article/pii/S2666691X20300142>

WS-2: HomeSeeker/ A visual analytics system of real estate data

<https://www.sciencedirect.com/science/article/pii/S1045926X17301246>

WS-3; Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains-cellular automata urban growth model

<https://www.sciencedirect.com/science/article/pii/S0143622813000362>

Scott’s Lit Survey

|  |  |
| --- | --- |
| [SB-1] | Rubén Hernández-Murillo, Michael T. Owyang, and Margarita Rubio. 2017. Clustered housing cycles. *Reg. Sci. Urban Econ.* 66, (2017), 185–197. |
| [SB-2] | Massimo Cecchini, Ilaria Zambon, and Luca Salvati. 2019. Housing and the city: A spatial analysis of residential building activity and the Socio-demographic background in a Mediterranean city, 1990–2017. *Sustainability* 11, 2 (2019), 375. |
| [SB-3] | Velma Johnson. 2018. A Spatial Analysis of Red X Properties and its Correlation to Foreclosed Properties within the City of Chicago. *A Thesis Submitted to the Faculty in the Department of Geography in Partial Fulfillment for the degree of Masters of Geography*, Chicago State University (2018). |
| [SB-4] | Melissa Shakro. 2013. Tracking neighborhood development and behavioral trends with building permits in Austin, Texas. *J. Maps* 9, 2 (2013), 189–197. |
| [SB-5] | Margherita Carlucci, Efstathios Grigoriadis, Giuseppe Venanzoni, and Luca Salvati. 2018. Crisis-driven changes in construction patterns: evidence from building permits in a Mediterranean city. *Hous. Stud.* 33, 8 (2018), 1151–1174. |
| [SB-6] | Arjun Subramanyam Varalakshmi, Chong Wang, and Christoph F. Eick. 2019. Fast proximity graph generation with spark. In *Proceedings of the 8th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, ACM, New York, NY, USA. |
| [SB-7] | Zhiliang Zhu, Jie Liang, Deyang Li, Hai Yu, and Guoqi Liu. 2019. Hot topic detection based on a refined TF-IDF algorithm. *IEEE Access* 7, (2019), 1–1. |
| [SB-8] | Isabella Gagliardi and Maria Teresa Artese. 2020. Semantic unsupervised automatic keyphrases extraction by integrating word embedding with clustering methods. Multimodal technol. interact. 4, 2 (2020), 30. |

**Brian’s Literature Survey**

**Now-Casting Building Permits with Google Trends**

Coble, David and Pincheira, Pablo M., Now-Casting Building Permits with Google Trends (February 1, 2017). Available at SSRN: <https://ssrn.com/abstract=2910165> or [http://dx.doi.org/10.2139/ssrn.2910165](https://dx.doi.org/10.2139/ssrn.2910165)

The researchers propose a forecasting model using up-to-date construction search terms from Google Trends in order to supplement the lagging permit indicator. Our forecasting model is also subject to this data lag and we could supplement our forecasting model with Google search terms as the researchers did here. The researchers see value in combining factors for forecasting, which they stop short of pursuing. In our project, we would combine Google Trend forecasting *for a specific region* with the corresponding permit data.

**Univariate and Multivariate Arima Versus Vector Autoregression Forecasting**

Bagshaw, Michael L., 1987. “Comparison of Univariate ARIMA, Multivariate ARIMA and Vector Autoregression Forecasting,” Federal Reserve Bank of Cleveland, Working Paper no. 86-02.

In this paper, Bagshaw compares the performance of 4 forecasting models: ARIMA, MARIMA, VAR, and BVAR on general economic data (EG unemployment rate and GNP). Bagshaw demonstrated that MARIMA generated the most accurate forecasts. In our project, we will use a time-series forecasting model and this paper serves as a foray into several popular models which we could implement and select the best on for final analyses. Bagshaw compares several models, but he does so all on the same data set. Our analysis expands on this shortcoming by implementing the same models on a different category of data – building permits.

**The Other Side of the Broken Window: A Methodology that Translates Building Permits into an Ecometric of Investment by Community Members**

O’Brien, D.T., Montgomery, B.W. The Other Side of the Broken Window: A Methodology that Translates Building Permits into an Ecometric of Investment by Community Members. *Am J Community Psychol* **55,** 25–36 (2015). https://doi.org/10.1007/s10464-014-9685-8

This paper proposes a methodology for assessing health based on infrastructural investment sourced from building permits. The researchers’ data is nearly the same as ours and they establish conventions and processing methodologies that we could use to prepare our data set. Given the nebulous nature of the subject, the researchers fail to establish a causal relationship between their indicators and community health. Our project is unlikely to establish one either, but we hope to add another perspective from which to view such a relationship.