

Project WASP: Time series Analysis of Construction Permits.

Problem Statement: Currently there are no free tools or services that combine geo-spatial and temporal analysis of construction data. Such an analysis, extended to factors beyond pricing, like density will greatly reduce decision making time and add value to our users.

Heilmeier Questions:

- A. **What are you trying to do?** - *Our aim is to build an interactive portal to visualize construction density trends in an urban area and perform time series analysis.*
- B. **How is it done today; what are the limits of current practice?**
1. *Lot of static visualization websites based on either geo-spatial or temporal but not both.*
 2. *Existing analysis are mainly focused on prices than density of constructions.*
- C. **What's new in your approach? Why will it be successful?**
1. *Dynamic and Interactive visualization of geo-spatial and temporal analysis.*
 2. *Focused on construction densities to view how markets and people are moving.*
- D. **Who cares?**
1. *Real Estate Analysts – Project the growth based on density movement.*
 2. *City Planners – Make data driven decisions for new infrastructure development.*
 3. *Retail Users - Get a holistic view of upcoming development in an area.*
- E. **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.)?**
1. *This tool will reduce decision making time of stakeholders by aggregating construction trends and time-based analysis.*
 2. *The tool's adoption can be assessed by measuring the traffic to the webpage (with Google Analytics). User surveys can be conducted to prove the hypothesis about the tool.*
- F. **What are the risks and payoffs?**
- Risks:** -Data cleaning can get complex and time consuming.
-Geo-Spatial visualization using libraries can become risky if they don't support our scope.
-Scope of the project might be difficult to manage based on time spent on homework.
- Payoffs:** - The ability for retail users to make sound financial decisions.
- City planning can use this tool for data driven decisions vs anecdotal instincts.
- G. **How much will it cost?**
1. *Resource Cost (10 hrs/person/week) = 50 hrs/wk * 8 weeks = 400 resource hours*
 2. *Storage Cost & Computation- 26\$/month(Tier 2= 1\$/month, AWS ec2 pricing of 0.0139/hr)*
 3. *Miscellaneous – 1% of the overall cost.*
- H. **How long will it take?**
7 weeks for project to be completed, 1 week for presentation prep.
- I. **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	Btran411, aparwal7	03/26/2021
POC for visualization with fixed data	Pkubsad3, wsultan3, dbader7	04/09/2021
Time series analysis of data	All	04/15/2021
Final product with analysis	All	04/25/2021
Final Presentation and wrap up	All	05/01/2021

Literature Survey:

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. Lot of work has been done on usage choropleth maps to visualize geo spatial models, like dynamic increase in perceivable area[PK-2], boundary neighbor 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 first two papers provided us guidelines on what ML models to use to predict growth such as ARIMA, exponential smoothing [1] and mix of Markov chain and the Cellular Automata [2]. The last paper was more focused on types of visualization and their best of use [3]. Lastly, to overcome associated potential challenges related to complexity of both the visualization and modeling we plan to use third party platforms as a service.

The study[AP-1] mainly wants to address how construction permits for residential, commercial or public buildings correlate with socio-economic demography of an area. 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. With flexibility and cloud computing, we reduce such limitations. The study[AP-2] identifies damage and recovery efforts based on building permits and spatial scans. Our tool aims to enable city planners to balance giving out building permits by understanding disaster recovery clusters and allocate resources accordingly. The study[AP-3] looks to utilize density of population to dynamically adjust k value in the algorithm as even within city concentration of building permits needs to be changed for example, city center vs suburbs etc. Our analysis using the building permits will focus on clusters of building permits than treating each permit in a silo.

Previous research using construction data, show that it can yield meaningful insights, in terms of trends and event linkage [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., MS Excel) 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 interactivity and better practices (avoid red-green color schemes. [SB-2]. We consider various techniques for analyzing [SB-2] [SB-3] [SB-6] our construction data set. We consider joining our base data with complementary data, based on news articles, which we could harvest either by web-scraping or 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.

[BT-1]The researchers propose forecasting with construction terms from Google Trends. Our forecasting model is subject to data lag and we could supplement our forecasting model with search terms similarly. The researchers stop short of combining predictors. In our project, we would combine Google Trend forecasting with the corresponding permit data. [BT-2] Bagshaw compares 4 forecasting models. In our project, we will use a TS forecasting model. This paper serves as a foray into several popular models. Bagshaw compares his models on the same data set. Our analysis expands on this by implementing the same models on construction data. This paper[BT-3] proposes a methodology for assessing community health based on infrastructural investment. The researchers establish data processing conventions we could on our data set. The researchers fail to establish a causal relationship. Our project is unlikely to establish one either, but we hope to add another perspective from which to view such a relationship.

References:

Prashant Kubsad's Literature Survey Source

[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]	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

Wael Sultan's Literature Survey Source

[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

Aayush Parwal's Literature Survey Source

[AP-1]	The Future of Spatial Analysis in the Social Sciences https://www.tandfonline.com/doi/abs/10.1080/10824009909480516
[AP-2]	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
[AP-3]	Adaptive clustering algorithm based on kNN and density https://www.sciencedirect.com/science/article/pii/S0167865518300266

Scott's Literature Survey

[SB-1]	Rubén Hernández-Murillo, Michael T. Owyang, and Margarita Rubio. 2017. Clustered housing cycles. <i>Reg. Sci. Urban Econ.</i> 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. <i>Sustainability</i> 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. <i>A Thesis Submitted to the Faculty in the Department of Geography in Partial Fulfillment for the degree of Masters of Geography</i> , Chicago State University (2018).
[SB-4]	Melissa Shakro. 2013. Tracking neighborhood development and behavioral trends with building permits in Austin, Texas. <i>J. Maps</i> 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. <i>Hous. Stud.</i> 33, 8 (2018), 1151–1174.
[SB-6]	Arjun Subramanyam Varalakshmi, Chong Wang, and Christoph F. Eick. 2019. Fast proximity graph generation with spark. In <i>Proceedings of the 8th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data</i> , 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. <i>IEEE Access</i> 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. <i>Multimodal technol. interact.</i> 4, 2 (2020), 30.

Brian Tran’s Literature Survey:

[BT-1]	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
[BT-2]	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.
[BT-3]	The Other Side of the Broken Window: A Methodology that Translates Building Permits into an Econometric 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 Econometric of Investment by Community Members. <i>Am J Community Psychol</i> 55, 25–36 (2015). https://doi.org/10.1007/s10464-014-9685-8