FRIEDRICH NAUMANN FOUNDATION For Freedom.
Pakistan

Consumption Patterns in Urban Pakistan

An exercise in data synthesis

White Paper

IMPRINT

Office

Friedrich Naumann Foundation for Freedom- Pakistan
P.O. Box 1733, Islamabad
Website: www.freiheit.org/pakistan
Facebook : @FNFPakistan
Twitter: @FNFPakistan

Authors

Hisham Bin Sajid, Director/Co-founder, Karachi Futures
Syed Ali Abidi, Associate Econometrician, Karachi Futures

Layout

Rebea Firdous, Communication Manager, FNF Pakistan

Editorial Staff

Aniqa Arshad, Program Manager, FNF Pakistan
Puruesh Chaudhary, Founder and President, AGAHI

Contact Info

Ph: + 92 (51) 26 55 750
Fax: + 92 (51) 26 55 752

Year

August 2022

Information on the use of this Publication

This publication is an information offer of the Friedrich Naumann Foundation for Freedom based on the research conducted by the Karachi Futures (KF) team (commissioned in 2021). It is available free of charge and not intended for sale. It may not be used by parties or election workers for the purpose of election advertising during election campaigns (federal, state or local government elections, or European Parliament elections).

Disclaimer

Every effort has been made to ensure the accuracy of the contents of this publication. The authors or the organization do not accept any responsibility of any omission as it is not deliberate. Nevertheless, we will appreciate provision of accurate information to improve our work. The views expressed in this white paper do not necessarily represent the views of the Friedrich Naumann Foundation for Freedom.

CONTENTS

LIST OF FIGURES

FNF PAKISTAN

FNF Pakistan Office has been working for a peaceful and progressive Pakistan since 1986. As individuals, everyone is entitled to personal dignity and freedom. It has been FNF's mission to make these principles valid for all. FNF promotes a social and political environment where every individual can assume responsibility for their life.

To know more about FNF's work visit
www.freiheit.org/pakistan

AGAHI

AGAHI Enterprise Private Limited is a professional services firm, and a counsellor to businesses and institutions. Karachi Futures is supported by AGAHI.

Karachi Futures is community-led research group and tech start-up that conducts research and builds tools on the intersection of public policy, strategic foresight, and data science.

KARACHI FUTURES

karachifutures.com

This paper aims to contribute towards improvement, efficiency and effectiveness of public service delivery and poverty alleviation at the local government level. This document provides a novel method to calculate consumption at a fine geospatial scale in Urban Pakistan, by synthesizing and transforming several different open-source datasets, and passing them through a statistical pipeline that generates powerful conclusions from relatively sparse data.

INTRODUCTION

A recurring issue when studying any kind of consumption patterns in Pakistan is the lack of concrete, well put together data (Elahi., 2008). This is in stark contrast to the amount of datasets available, of which there are many. The main issue, when working with Pakistani, or indeed South Asian data in general, is the need to synthesize data sources through means beyond just standard joins. In particular, much individual-level information can be gleaned through a combination of large scale, open source data, and smaller scale surveys. In this paper, we demonstrate a method for estimating demand or consumption of specific goods at fine geospatial resolutions. We attempt to provide a baseline model which we hope will become the basis of a much more complex model for understanding consumption patterns in urban centres in Pakistan at a much more granular level.

"THE USUAL ISSUE WITH WORKING ON DATA-DRIVEN PUBLIC POLICY IN PAKISTAN IS THE LACK OF EXISTENCE OF CLEAN AND USEFUL DATA SETS"

SIGNIFICANCE

Previous research on economic activity and consumption has been done through data collected through targeted surveys, questionnaires, and field studies conducted either in person or via Telephone calls or emails. This method, though relatively thorough and effective can be prohibitively expensive. The cost inhibits institutions from collecting a lot of data at a higher frequency, particularly in poorer parts of the world.

However, the advent of big data i.e. large quantities of data being collected at a higher frequency, often through automated methods, has opened up the possibility of exploring new methods in economic analysis (Harding, 2018). This trend has been quickly adopted by leading research institutes across the world. For example, the MIT Senseable City Lab (SENSEable, n.d.) has done many research projects along the intersection of economic analysis and big data.

In one of their projects, it is demonstrated that local restaurants can be used to reasonably estimate spatial distribution of socioeconomic activities at a fine geospatial resolution. The research involves using open source data collected from Dianping, the largest online rating and deal service platform in China (similar to Yelp in the US, and in some ways Foodpanda in Pakistan), to predict four important socioeconomic variables: daytime population, night-time population, number of firms, and volume of consumption at various spatial resolutions (Dong, et al., 2019). The paper attempts to solve a problem that is very much similar to ours, but by using restaurant data. In another study, anonymized debit and credit card transaction data was sourced from one of Spain's largest banks, the Banco Bilbao Vizcaya Argentaria (BBVA). Using data from 2011, for 4.5 million active customers executing well over 178 million transactions, researchers were able to cluster cities in Spain using people's spending habits and behaviours (MIT, n.d.). They were also able to study how spending habits were effected by age and gender, with women carrying out more transactions then men, but men carrying out more transactions farther (distance-wise) from their home address (Sobolevsky, et al., 2016). Although not very practical in Pakistan due to low credit and debit adaption in country (Khalid, et al., 2013), it gives us interesting insight into how big data can be used to analyse urban areas at scale.

There have been previous studies conducted in the developing world that have attempted to use statistical estimation techniques to overcome the lack of well-documented data. One example is from Bangladesh (Rahman, et al., 2020) where Bayesian techniques were used to aggregate expenditure on a particular good by geographic units. Which uses a more straightforward Bayesian regression to estimate the costs of healthcare expenditure for each unit, and aggregate upwards. This ease comes from their geocoded data. Our challenge is similar to theirs, except we also lack data that is geocoded.

The data used here is a cross-sectional survey of 1593 randomly selected households was carried out in Bangladesh (urban area of Rajshahi city), in 2011. Our study aims to recreate these results more generally, i.e. for more than just healthcare expenditures.

Similar to the Healthcare expenditure paper above, Peter M. et al (Haas, et al., 2008) looks into Transportation costs by characteristics for specific neighbourhoods in the US. They attempt to create a statistical model to predict household total annual transportation expenditures for each neighbourhood in the largest metropolitan regions in the United States, controlling for the economic environment and household size and income. Transport is a bit trickier than healthcare, since it's exceedingly dependent on the environment of the household. We do not have the data to do a thorough analysis of estimating transport costs, however we are able to use self-reported measures of expenditure on transport in our study.

"IN ORDER FOR OUR MODEL TO BE REPLICABLE ACROSS MULTIPLE CITIES IN PAKISTAN AND TO BE COST EFFECTIVE, WE FOCUSED ON OPEN AND PUBLICLY AVAILABLE DATA"

DATA

The usual issue with working on data-driven public policy in Pakistan is the lack of existence of clean and useful data sets. There is no centralized data portal where you can extract relatively clean flat files to then put into your models. One example of a portal of this nature is Data.gov (Lakhani, et al., 2002) , which is branded as "The home of the US Government's open data" and was launched by the US Government in 2009 to increase access to high value machine readable data sets. The closest comparison to a similar initiative in Pakistan is the Data4Pakistan (Ali & Imran, 2020) tool developed under the Ehsaas program in partnership with the World Bank Group. The portal has access to 120 development and policy indicators at the district level along with poverty estimates.

In a post-pandemic world, having granular data updated frequently has become essential for policy makers to make quick and effective decisions (Haldane & Chowla, 2021). Data frequency is currently out of scope of this paper, and our focus is mostly on generating data at a finer spatial resolution. Furthermore, when sourcing datasets that could be used in our model we were able to enumerate the following data type by ease of access, cost to acquire, and completeness.

A. Open and publicly available data consumable as either flat files or via some public access API: One example of such data sources includes the Open Street Maps (OSM), and its Overpass Turbo API. Another example would be data we downloaded from portals like the UN Humanitarian Data Exchange, World Bank Open Data, and Gapminder (Lang, 2012).

B. Open and publicly available data that requires pre-processing / cleansing before use: Examples include data available with Federal Board of Revenue (FBR) in pdf files, Household Integrated Economic Survey (HIES) microdata with the Pakistan Bureau of Statistics (PBS) available as STATA files that need to be manually decoded, web data that needs to be scraped from an open-access website and later transformed for use. A subset of this data also includes data in raster format available on platforms like Google Earth Engine that requires significant pre-processing before it can be used.

C. Closed access data that can be accessed via some standard paid API: The best example of this is data available via Google APIs, specifically for our use case, APIs that fall under the Google Maps Platform which include but are not limited to Directions API, Places API, Geocoding API.

D. Closed data in silos: This would include government data that is not publicly available, or data generated by corporations that can be useful for our model. For example, anonymized geospatial customer data generated by large telecommunication corporations like Telenor, Jazz etc. or anonymous retailer data generated by large Consumer Packaged Goods (CPGs) companies like Unilever, Nestle, Reckitt.

In order for our model to be replicable across multiple cities in Pakistan and to be cost effective, we focused on open and publicly available data. We shall see in the following sections the data sets we are currently using, why we are using them, and how they are being used. Following is summary followed by detailed descriptions of data sources used:

Table 1: Summary of Data sets used in this model

Data Set	Description	Date
Population Density Map by Facebook (Meta) Data for Good	Population data along with gender and age cuts.	March 2021
FBR Immovable assets data table	Per marla* value of commercial and residential property as estimated by the Federal Board of Revenue.	February 2019
Household Integrated Economic Survey (HIES)	Household survey data collected at district level by Pakistan Bureau of Statistics (PBS)	2018

* 1 Marla = 30.25 square yard

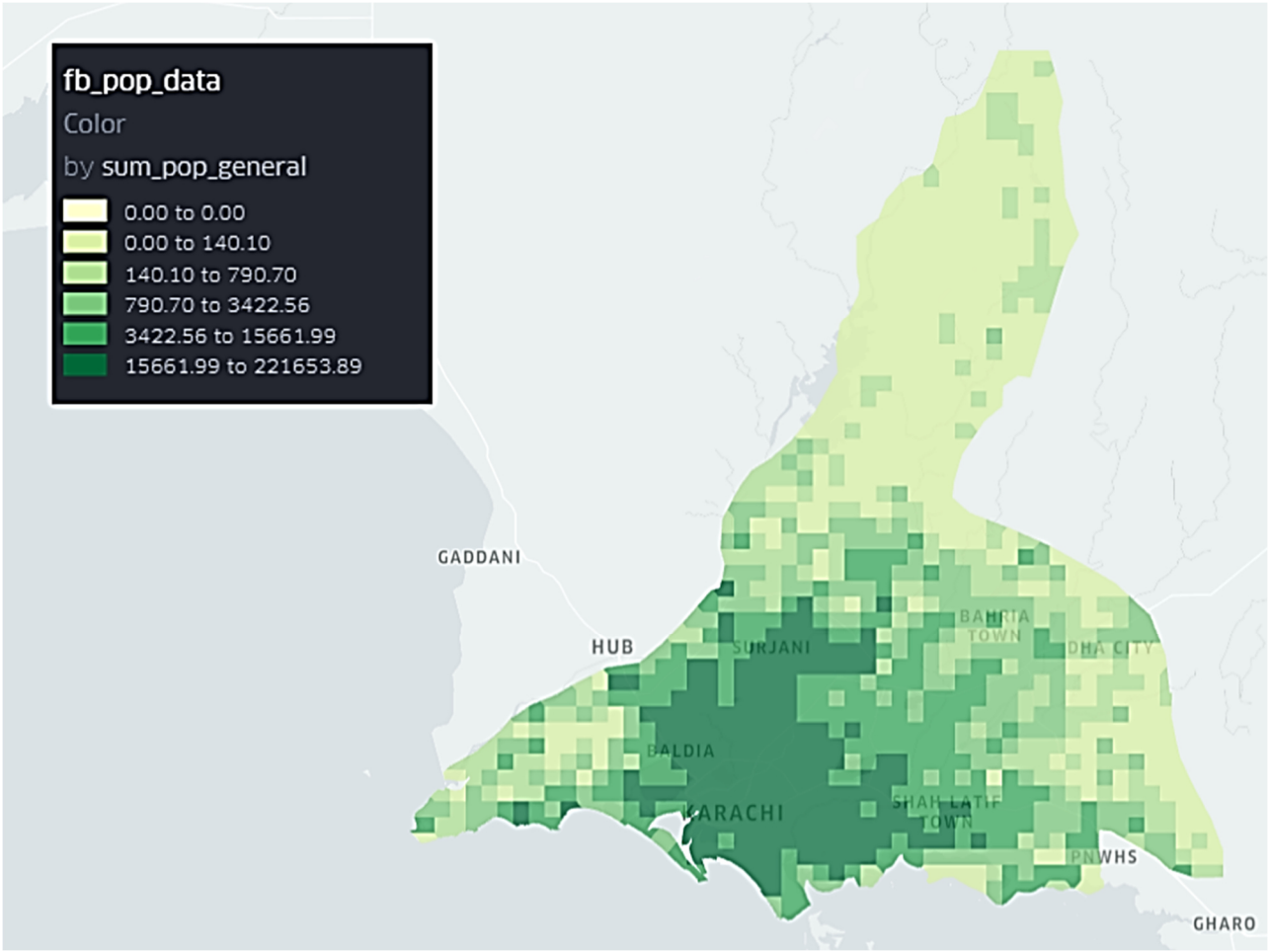
POPULATIOPN DATA - FACEBOOK DATA FOR GOOD

There are multiple sources for granular raster population data available at multiple geospatial resolutions. Some globally recognized open population datasets that have been used for some time now include Worldpop constrained and unconstrained population of the world data that is updated annually, and is available at a 100x100m granularity. Another source is the Gridded Population of the World (Doxsey-Whitfield, et al., 2015) dataset which is a product of NASA's Socioeconomic Data and Applications centre (SEDAC) hosted at Columbia University which is also available at 100x100m granularity.

For our model we choose to work with high resolution population density maps created by Facebook Data for Good (Tiecke, et al., 2017). This was primarily due to the fact that the data was available at a much finer spatial resolution of 30x30m and we had gender and demographics data available as well; including variables like number of men, women, children, youth, elderly and women of reproductive age, some which as we will see are being employed in our model.

The population data available for Pakistan is dated for March 2021.

Figure 1: Population Heat Map for Karachi



FEDERAL BOARD OF REVENUE (FBR) PROPERTY VALUE

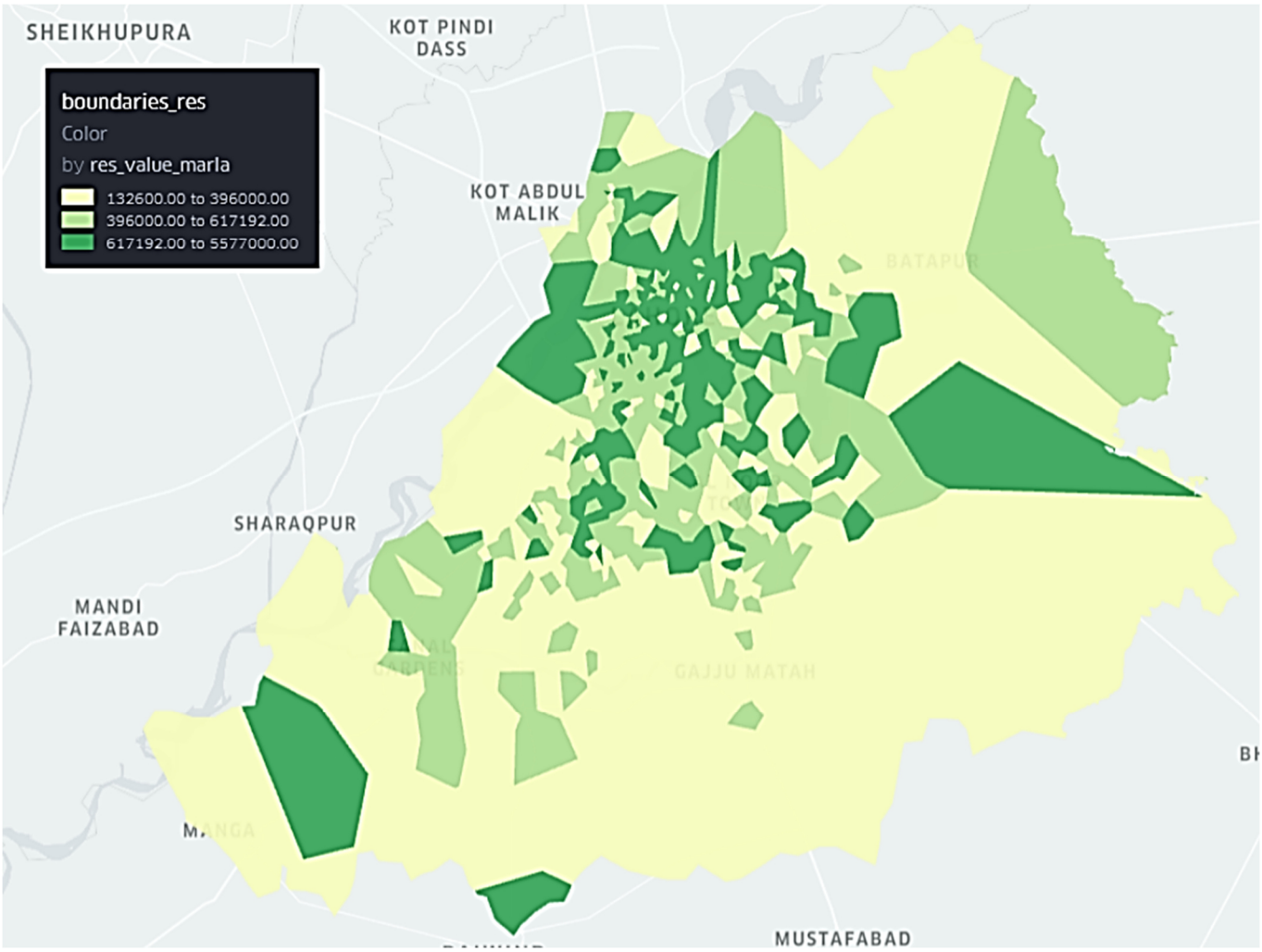
In order to make our model work we needed some proxy of how cheap or expensive an area is to live in, for which we explored multiple open source data sources and proxy variables. Eventually we settled on Valuation of Immovable Properties conducted and compiled by the FBR for major cities across Pakistan. This data is made available to the general public through the Federal Board of Revenue's website (FBR).

Although the data format and the variables vary slightly from city to city and province to province, for almost all cities we have the value of residential and commercial property by locality.

The values from the FBR data tables might not reflect the actual value of residential or commercial value in an area and have recently come under fire for the same reason (Tribune, Dec 2021) - with claims that the prices are lower than market value since that results in lower amount of payable taxes (Niazi, Dec 2021) . However, they do give us a directional idea. For example, the per Marla value of land in a less developed area of Lahore like Mandian Wala will be much less than the per Marla value of land in Bahria Town Lahore.

FBR Immovable property data is dated 1st February 2019, as per SRO120(1)/2019.

Figure 2: Property Value for Lahore



HOUSEHOLD INTEGRATED ECONOMIC SURVEY (HIES)

The Pakistan Social and Living Standards Measurement Survey(PSLM) is a regular survey conducted by the Pakistan Bureau of Statistics since 1st July 2015. It is specially designed to provide social and economic indicators at the household level for policymakers in alternating years at the district and sub-district level. The data generated through these surveys is an integral part of the government's anti-poverty measures, which in-turn play an important role in achieving the United Nation's Sustainable Development Goals, specially SDG1; No poverty. There are two versions of the PSLM, one of each at the district and provincial level. For our study, we utilize the district level survey. The total sample of the survey consists of 80,000 households.

We use a version of the HIES dataset from 2018, which is the latest version of the survey for which micro-data is publicly available. The data contains a wide range of variables.

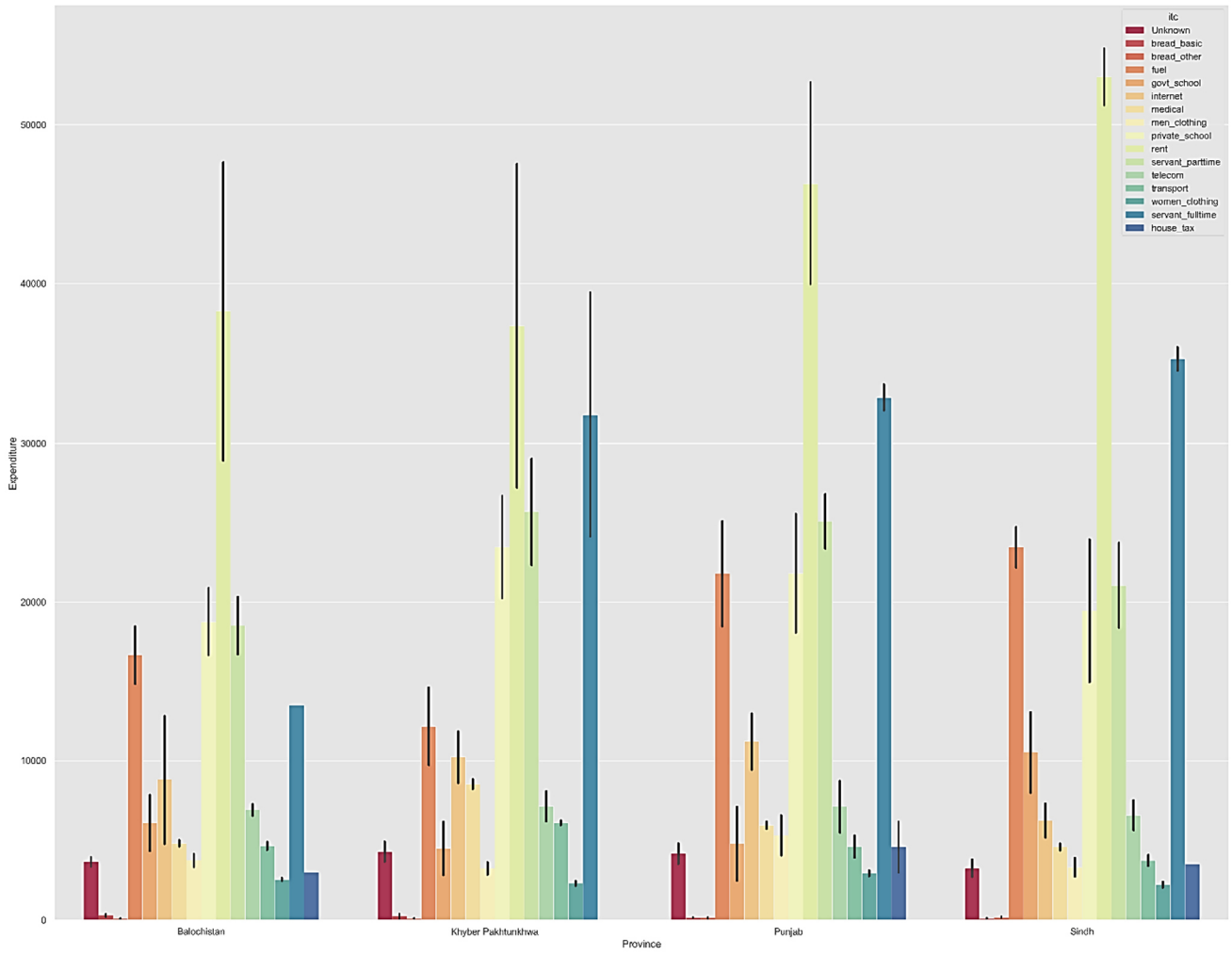
Of particular interest to us are the following:

Expenditure on the following:

- Food, including *Roti*(bread)and luxury foods.
- Clothing, for both men and women.
- Rent, as well as housing tax.

The strength of our result lies on the conclusion that with just these variables, and our collective demographic data we are able to map out expenditures across our three cities of choice in Urban Pakistan (and as many more as is required).

Figure 3: Expenditure Distribution in HIES



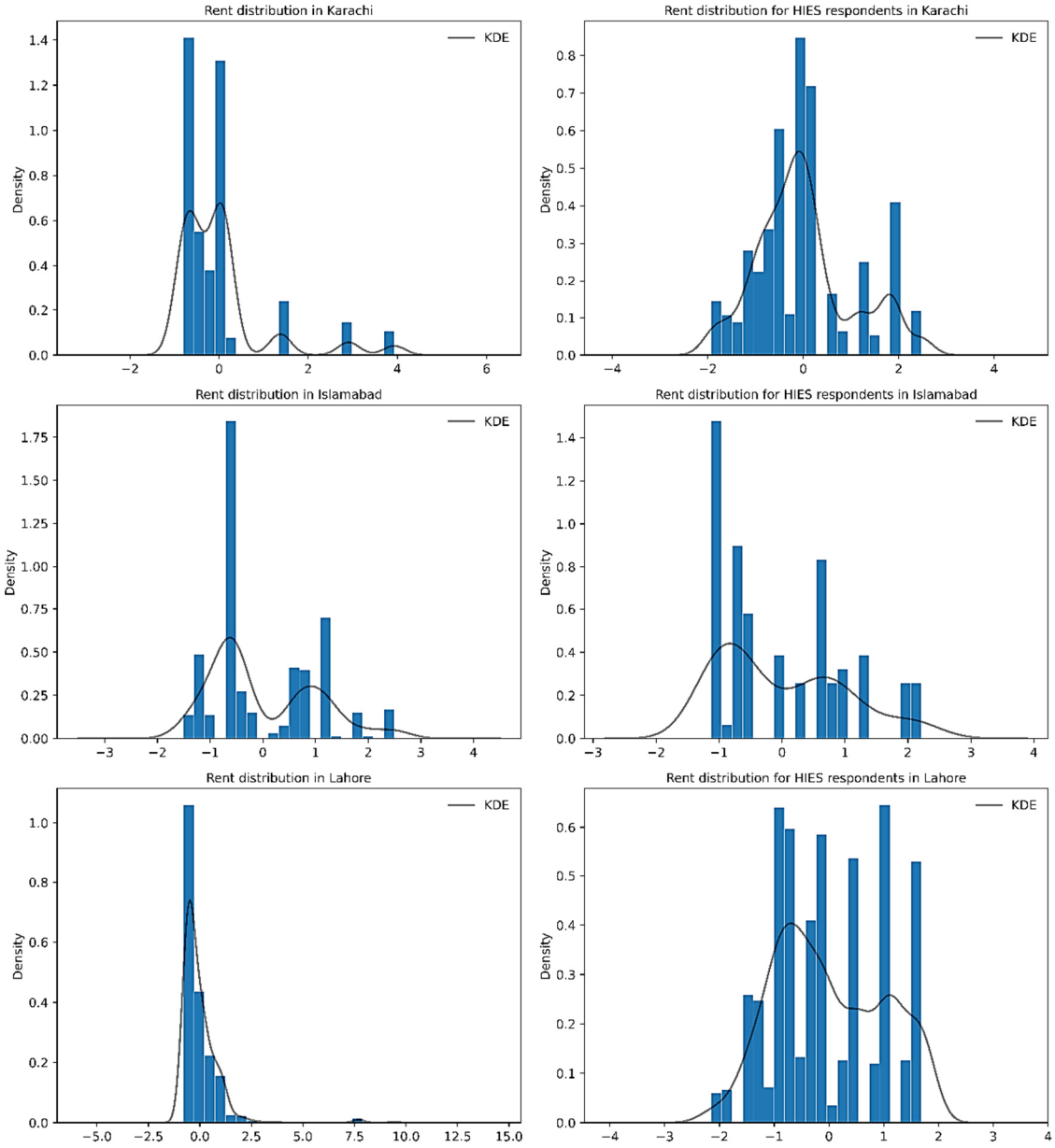
THEORATICAL MODEL

BASIC SETUP

Our data universe consists of N individuals each for a number of cities in Urban Pakistan (Karachi, Islamabad and Lahore in our specific case). We divide each city into M blocks (in our case 1 Km by 1 Km). Each of these M blocks is further subdivided into K sub-blocks, which are significantly smaller, and are indexed by k.

We aggregate our K blocks into M blocks due to dimensionality issues, and to have a distribution within each block for observable factors. For each individual, we have their annual expenditure in PKR in each of several categories X_i. Our objective is to attain a value of X_j for each of the j city blocks.

Figure 4: Distribution of rent values in HIES



MODEL

We model the expenditure of goods in each arbitrarily small city block using the following framework:

$$X_j = \sum_{i=1}^{N} X_i p_{i,j} w_i \tag{1}$$

where N is the total number of individuals represented for a city in our HIES dataset, and $w_{i,j}$ is the sample weight presented in the HIES. A detailed breakdown of the calculation of these weights is presented on the HIES website PSLM-HIES. The crucial object of interest, and the subject of our modelling is the $p_{i,j}$, which is the probability of an individual i being present in quadrant j.

We rely on a few assumptions, backed by the data we observe. They are as follows.

A. **Rent values and Rent expenditures are roughly normally distributed.**

This is our most demanding assumption, but one we see panning out in the data. As can be observed in the plots above, the kernel density estimates of all three cities match a roughly normal distribution. This is the distribution of rent across cell blocks. Since our cell blocks are arbitrarily sized, and we can imagine each block consisting of smaller blocks, it is no stretch to assume rent distributions within a particular block are distributed normally, but with an unknown mean and variance. Part of our exercise is a method to find these parameters.

B. **The variance in population sizes within cell blocks is directly proportional to the variance of rent**

We use this assumption, but rely on regression results to do so and do not make assumptions about the direction of this relation. It is purely driven from data. With these assumptions and data in place, we model $p_{i,j}$ as draws from the following distribution:

$$p_{i,j} \sim N(\Omega_j, \Sigma_j) \tag{2}$$

Where Ω_j , Σ_j are the a mean and variance matrix measure 2x1 and 2x2 respectively. They can be broken down into their components as follows:

$$\Omega_{i,j} = \begin{pmatrix} L_j \\ R_j \end{pmatrix} \tag{3}$$

And

$$\Sigma_j = \begin{pmatrix} \sigma_{L,j} & 0 \\ 0 & \sigma_{R,j} \end{pmatrix} \tag{4}$$

Each of these objects are modelled separately, and used as the basis of a multivariate normal distribution. Broadly, we have two factors in the mean matrix. The first is a function of the average age mix in the cell block j. The second is a function of the cell block j's rental costs, as well as the variance in population within the block. Formally:

$$L_j = \sum_{k=1}^{K} Age_k \tag{5}$$

Where Age_k is the average age observed in sub-block. k Essentially, we are just taking the sample mean. R_j is determined in a more straightforward way from the data in the FBR valuation tables.

$$\sigma_{L,j} = \varphi_j \tag{6}$$

$$\sigma_{R,j} = (\overline{Pop_k} - \underline{Pop_k})\epsilon_R \tag{7}$$

where Pop_k represents the maximum/minimum population within the κ cells in each cell-block. Essentially, the differential gives us a measure of variance of population within the cells. φ_j is the normalized standard deviation of population within the cells. This normalization renders each

value of φ between 0 and 1. Lastly, ϵ is a parameter given to us by the following estimation equation:

$$Rent_j = \beta + \epsilon_L L_j + \epsilon_R Pop_j + e \tag{8}$$

Essentially regressing youth population and overall population onto Rent. Hence we make no A priori judgements about the relationship between population density and rent. With this distribution in place, we calculate the probability of an individual living in cell block j as a function of individual i's age and rent expenditure, using the difference in cumulative density functions over a small range.

$$p_{i,j} = F_{N_j}(L_i + \delta_L, R_i + \delta_R) - F_{N_j}(L_i - \delta_L, R_i - \delta_R) \tag{9}$$

These values of delta are set to be small (0.05).

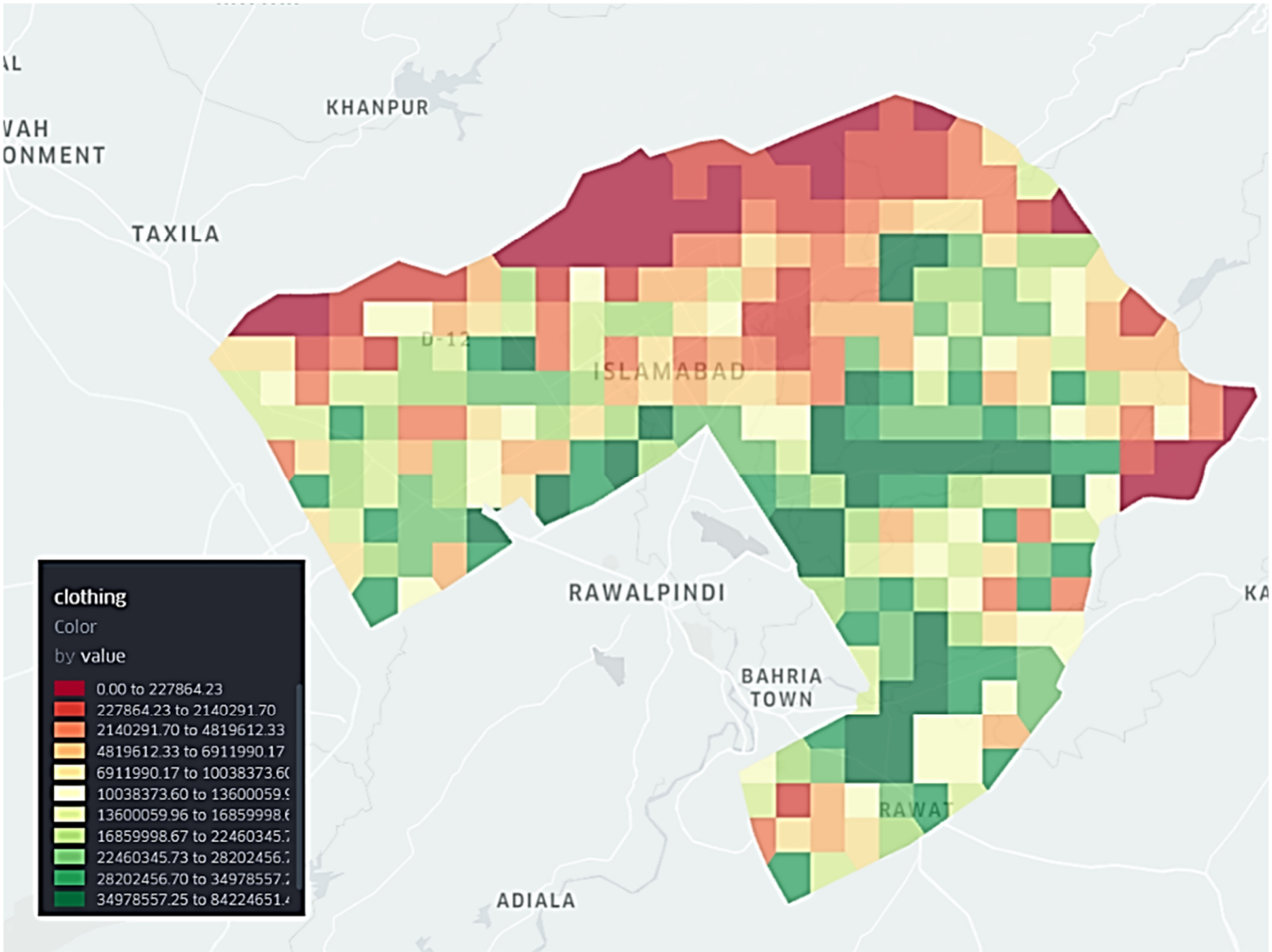
RESULTS

We calculate the expenditure by survey respondents on a particular good or commodity using equation 1.

For demonstrative purposes, we pick the amount of money being spent on women's clothing by each household. We run our model for Islamabad, Karachi, and Lahore divided into a grid of roughly 1 square kilometer cell. Following are some examples from the cities mentioned that add credibility to the results of our model.

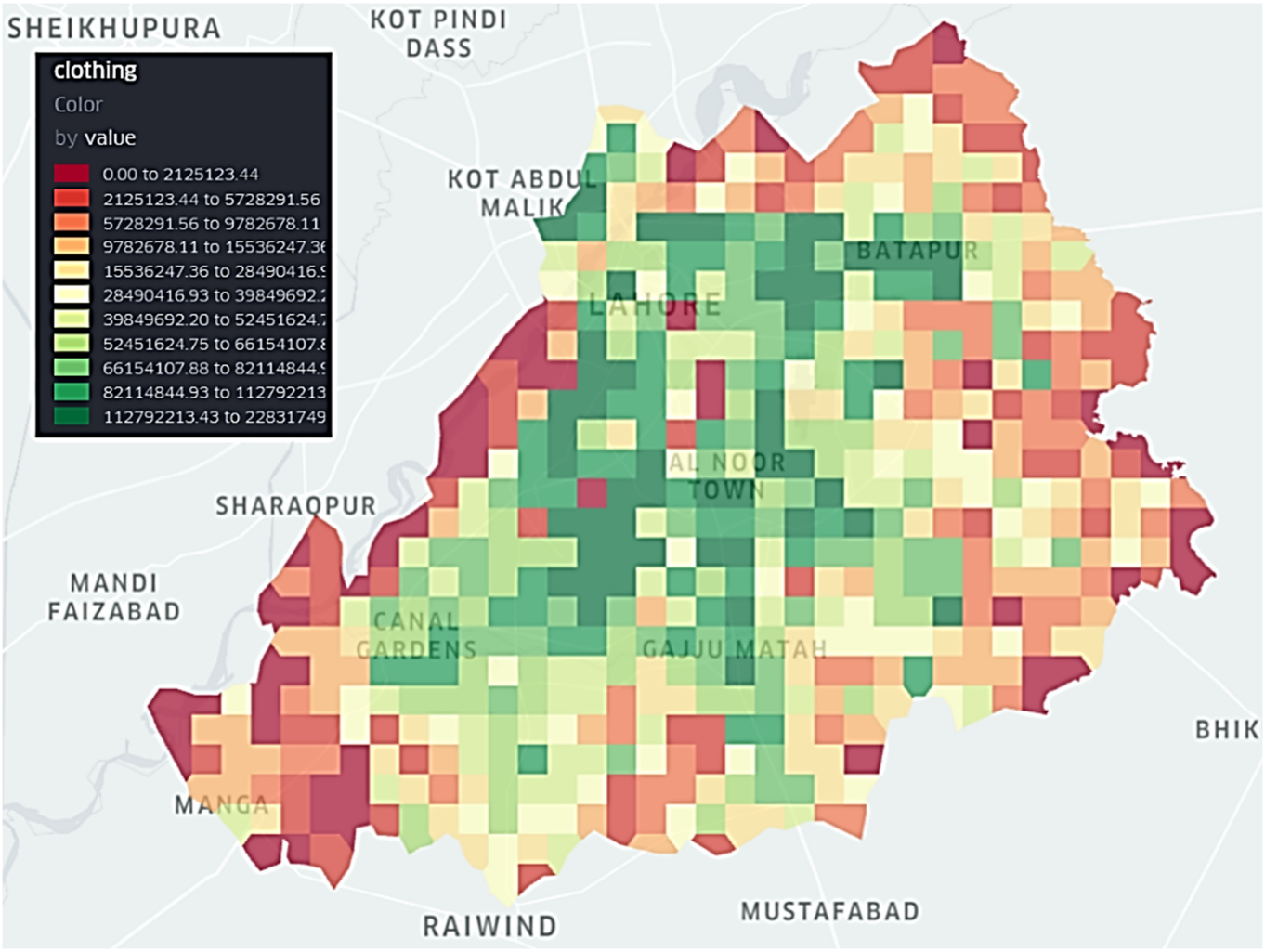
In Islamabad, if we look at the blocks that roughly make up I-8 and I-9 we see that residents in these areas spent roughly 70 million PKR on women's clothing, whereas if we look up the blocks that constitute the area of Chak Shahzad and Lakhwal, we see that that this amount sums up to around 10 million PKR. We can see the overall map of Islamabad in the figure below.

Figure 5: Expenditure on women's clothing for Islamabad



In Lahore, if we pick the most built up 1 square kilometer block from Bahria Town, we see that it spent 112 million PKR on women's clothing. If we pick a similar 1 square kilometer block around Mughal Pura we see that around 80 million PKR was spent on women's clothing. We can also see that parts of old Lahore due the sheer number of people living in these area show a very high cumulative spend. For example, a 1 square kilometer block around the locality of Krishan Nigar is spending 177 million PKR on the same commodity as per our model.

Figure 6: Expenditure on women's clothing for Lahore

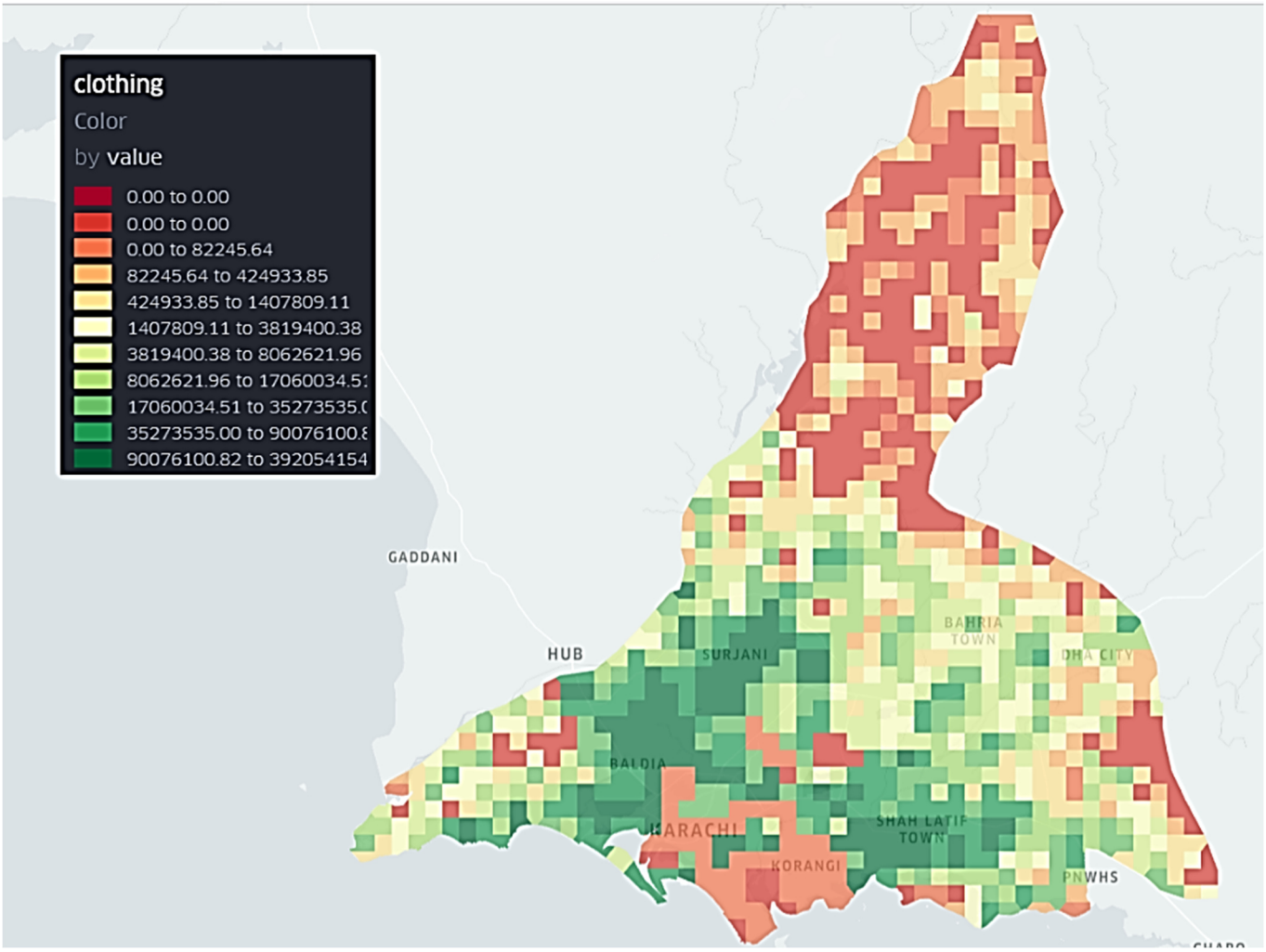


LIMITATIONS

There are however limitations with our approach, which are prominently visible when looking at the data for Karachi. We see a lot of areas where we'd expect high consumption levels, but are seeing very little to no consumption according to our model. This is primarily caused by not having finer property evaluation data. Since our model relies on variation of property values and demographics, it enormously deflates the consumption values in areas with low variation. A prime example of this phenomenon is the case for DHA Karachi. For all phases of DHA (except DHA City) we see a flat residential value of 48000 PKR/square yard, which means there's 0 variance, which leads to our distribution for these cells being constant.

Hence, unless an individual's rent is exactly 48000 PKR (which has a near 0 probability of occurring), they would be assigned a 0 probability of living in this area. This is why our model shows that no money was spent on women's clothing in DHA.

Figure 7: Expenditure on women's clothing for Karachi

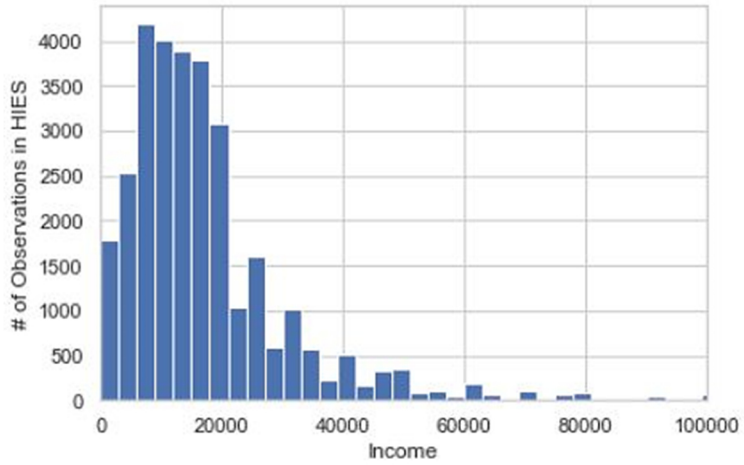


In addition, another limitation for our research comes from very sparse data on higher income individuals. This is demonstrated in Figure 8, where barely anyone in the sample earns more than PKR 80,000 per month. This is a problem for aggregating results in areas like Clifton, Karachi or the entirety of Islamabad, where high income individuals are the norm. The sparsity of samples means higher variance on the right tail of the distribution.

In a similar fashion we can look at other survey response data from the HIES at a fine geospatial resolution. We have run our model for multiple variables to understand consumption patterns in a more thorough manner, these variables include; men's clothing, bread, fuel, government and private school spending, transport, medical expenses, and internet expenses. Interactive maps for these variables are available in the GitHub repository (Jugnu-Github, n.d.) shared in the data appendix.

"SINCE OUR MODEL RELIES ON VARIATION OF PROPERTY VALUES & DEMOGRAPHICS, IT ENORMOUSLY DEFLATES THE CONSUMPTION VALUES IN AREAS WITH LOW VARIATION"

Figure 8: Income distribution in HIES



WAY FORWARD

In the results section we discussed how our model gives satisfactory directional results for some geographies, while for others the results could use a lot of improvement. We plan on improving the accuracy of our model by adding more data sources to substitute or augment our existing data. For example, when it comes to rent and property valuation, we need to look for other open data source which can be used as a proxy for how cheap or expensive an area is.

We also want to explore the use of other entirely different data sources to better tune our results; this includes multi-spectral satellite imagery to gauge things like urban build up and night light intensity, POI (Point of Interest) data sourced from either Google Maps or Open Street Maps (OSM, n.d.)

Using this data, and perhaps a more advanced modelling approach, we hope to build a definitive model (or models) for understanding consumption patterns in Pakistan. The resulting work will not only help policy makers and academics, but will also help corporations and start-ups grow in urban Pakistan in a more sustainable and data-driven way. There needs to be a focus on developing capacity and quality of the available healthcare systems in South Asia. Through technical and financial assistance, there needs to be focused development in the healthcare sector.

The developed countries need to help the weaker ones establish mechanisms to increase outreach to excluded or remote areas. There need to be tracking mechanisms in place which monitor the spread along with containment measures in all such areas. Additionally helping design effective awareness campaigns to communicate mitigation measures, support strategies and useful information.

“THIS WILL HELP POLICY MAKERS, ACADEMICS, CORPORATIONS & START-UPS TO GROW IN URBAN PAKISTAN IN A MORE SUSTAINABLE & DATA-DRIVEN WAY”

ANNEXURE A: DATA PREPERATION & CLEANING

HIES data - 2018

The HIES data is present in different formats, and is divided into sections. It is based on a long questionnaire, and different portions of that questionnaire are divided into different code files. Each row in each file is uniquely identified by a Household code, which can be decoded to know the area the household is present in, at the city level. We derive the columns we need from the HIES by merging the tables appropriately. The important component for us is expenditure data, and rent. The HIES also contains data on dwelling characteristics, such as numbers of rooms in the household's dwelling. We use this to normalize the rent, by dividing the reported rent of the household with the number of rooms to get a per room rent. These values of rent are then normalized by city, between 0 and 1, so they are on a purely ordinal scale.

Fishnet grids

Fundamental to our modelling technique is distributing data on to a 1x1 km grid. We create a fishnet grid for each city by creating a mesh of equally spaced points and bounded by the geometry of that city. The distance or space between each point of the mesh determines the area each grid cell of the fishnet grid will cover. Once the mesh is ready, we create Voronoi Tessellations (Moller, 2012) around each point of our mesh. We handle the infinite regions of our tessellations by intersecting them with a bounding geometry which is the boundary of the district or city in question. We create these fishnet grids for each city we want to run our analysis for. Each grid cell has a unique identifier which can be used for linking multiple data points to each cell.

FBR property data Feb, 2019

We can download FBR property data tables for different districts from the official FBR website. This data is available in PDF format; hence we wrote a pdf parser that converted these into flat files in csv format. After some manual cleaning, we have the name of each locality with the per Marla residential and commercial value. In order to add a geospatial element to this data, we geocode the localities and addresses after some basic pre-processing. This gives us point data for each locality, however, we still don't know the boundaries of these areas, and within the current scope of this research it was not possible to manually draw boundaries for each locality. For Lahore alone, we had data for almost 1200 neighbourhoods/towns/societies. We approximated the boundary of each locality by creating Voronoi Tessellations using the geocoded points. We can see this in effect in Figure 2.

We translate this data to our fishnet grids for each city by simply getting the centroid of each grid cell from our fishnet grid, and checking which locality boundary (extrapolated using the method described above) the centroid falls in. We assign the value of the locality to the grid cell to which the centroid belongs to.

Facebook Population Density Maps, March 2021

Facebook population density data can be downloaded in raster format from UN Humanitarian Data Exchange. The data is distributed into multiple files for different latitude and longitude, and demographic cuts. Our script picks the relevant file for the city in question, and clips the data to get a raster file for the city or district we want to work with. We convert this data into a more discrete vector format by computing the relevant demographics for each grid cell of our fishnet grid. We calculate the sum, mean, maximum, minimum and standard deviation for each grid cell. It is worth mentioning here that our density map has a 30x30m granularity and we are aggregating it to a 1x1 km grid.

GitHub Repository

We have created a GitHub that will house part of the code and all of the maps with the final output for all the HIES variables that we explored. This repository can be accessed via

https://github.com/hishamsajid/disaggurbanpak_kf/

REFERENCES

Ali, A. & Imran, M., 2020. The evolution of national spatial data infrastructure in Pakistan, implementation problems and the way forward.. *International Journal of Spatial Data Infrastructures Research.*

Dong, L., Ratti, C. & Zheng, S., 2019. Predicting neighborhoods' socioeconomic attributes. *Proceedings of the national academy of sciences,* 116(31), p. 15447–15452.

Doxsey-Whitfield, E. et al., 2015. Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4.. *Papers in Applied,* 1(3), p. 226–234.

Elahi., A., 2008. Challenges of data collection in developing countries–the pakistani experience as. *Statistical Journal of the IAOS,* 25(1,2), pp. 11-17.

FBR, n.d. *www.fbr.gov.pk.* [Online]
Available at: https://fbr.gov.pk/valuation-of-immovable-properties/51147/131220
[Accessed 8 June 2022].

Haas, P. M., Makarewicz, C., Benedict, A. & Bernstein, S., 2008. Estimating transportation costs by characteristics of neighborhood and household. *Transportation Research Record,* 2077(1), pp. 62-70.

Haldane, A. & Chowla, S., 2021. Fast economic indicators. *Nature Reviews Physics,* 3(2), pp. 68-69.

Harding, M. H., 2018. Big data in economics. *IZA World of Labor.*

Jugnu-Github, n.d. [Online]
Available at:
https://github.com/hishamsajid/disaggurbanpak_kf/

Khalid, J., Butt, H. S., Murtaza, M. & Khizar, U., 2013. Perceived barriers in the adoption & usage of credit cards in pakistan banking industry. *International Review of Management and Business Research,* p. 104.

Lakhani, K. R., Austin, R. D. & Yi, Y., 2002. Data.gov. *Harvard Business School Case 610-075.*

Lang, R., 2012. *Using Gap Minder, GW-Unterricht.* [Online]
Available at: https://www.gw-unterricht.at/images/pdf/gwu_126_076_087_lang.pdf

Moller, J., 2012. Lectures on random Voronoi tessellations. *Springer Science & Business Media,* Volume Volume 87.

Niazi, A., Dec 2021. *The Profit Pakistan Today.* [Online]
Available at:
https://profit.pakistantoday.com.pk/2021/12/05/the-fbr-real-estate-valuation-kerfuffle/
[Accessed 8 June 2022].

OSM, n.d. *Copy Right Information.* [Online]
Available at: https://www.openstreetmap.org/copyright

PSLM, n.d. *Pakistan Bureu of Statistics.* [Online]
Available at: https://www.pbs.gov.pk/content/microdata

Rahman, M. M. et al., 2020. Disease-specific out-of-pocket healthcare expenditure in urban bangladesh: A bayesian analysis. *PloS one,* 15(1), p. e0227565.

SENSEable, n.d. *Senseable City Lab.* [Online]
Available at: https://senseable.mit.edu/

Sobolevsky, S. et al., 2016. Cities through the prism of people's spending behavior. *PLoS ONE ,* 11(2), p. e0146291.

Tiecke, T. G. et al., 2017. Mapping the world population one building at a time.. *arXiv preprint,* p. arXiv:1712.05839.

Tribune, T. E., Dec 2021. *The Express Tribune.* [Online]
Available at: https://tribune.com.pk/story/2332776/fbr-directed-to-revoke-new-property-valuations
[Accessed 8 June 2022].

Urban-lens, S., n.d. *Urban Lens.* [Online]
Available at: http://senseable.mit.edu/urban-lens/
[Accessed 8 June 2022].

AUTHORS

Muhammad Hisham Bin Sajid, is a Data science and technology professional with extensive experience in product management, enterprise solution design, and data driven decision making. He has done BS in Computer Science from Institute of Business Administration (IBA) Karachi.

He is currently serving as Director/Co-founder, Karachi Futures.

Twitter: @hishamsajid
Email: hishamsajid113@gmail.com

Syed Ali Abidi is a practising Data Scientist working towards developing pairing algorithms, with years of experience in Academia both internationally and in Pakistan. He has done MA in Economics from McGill University Canada.

Currently he is serving as Associate Econometrician, Karachi Futures.

Email: mohammadaliabidi.94@gmail.com

END

Post Box 1733

Islamabad 44000 – Pakistan

@FNFPakistan