Theoritical analysis:

This provides annual data on 142 Indian cities for the years 2000–2011, spread over 47236 observations and 339 attributes under 5 primary categories: i. Demography ii. Education iii. Household iv. GDP/GVA v. Employment

The database is well structured and provides features in different units (for instance in the local currency and in USD), as well as variables related to Unemployment (such as inflation, mean household size, mean household expenditure, etc.) However, the dataset does not cover villages and towns, and hence rural India is missing. Also, since the data is aggregated at the city level, it does not provide demographics data at the household level (e.g. household income) or at the individual level (gender, educational qualifications, annual income, etc.) For this we turned to a second dataset.

**5th Employment Unemployment Survey (Labour Bureau, 2019)**

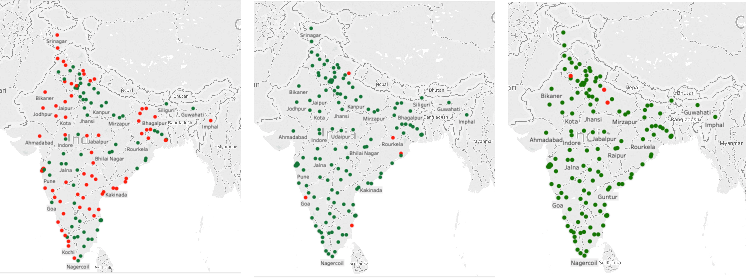
We obtained this dataset directly from the Ministry of Labour and Employment (Government of India). It pertains to the survey conducted by the Government between April to December 2015, across 36 States and Union Territories. The original survey covered over 700,000 individuals in over 150,000 households using a multi-stage stratified random sampling approach. This is more granular than the MarketLine dataset and covers the following categories: i. Age and demographics of individual ii. Training and education details iii. Duration of employment/unemployment iv. Reasons for unemployment

However, several survey questions used by the NSSO in earlier surveys have been dropped by the Labour Bureau (including religion, land ownership and marital status) which may have been important attributes for the purpose of our study. Also, we were not provided with the entire dataset, and around 200,000 observations have been withheld (details provided in the appendix).

We have used both the datasets in our study, for identifying the causes of unemployment amongst youth in India. Doing so, we have arrived at a business idea that will provide meaningful work of minimum 10 hours a week for 10,000 youth over the next decade.

Growth in employment in agriculture, industry and services reveals that most cities had a negative growth rate in the number of people employed in agriculture, while, employment in industry and services has grown in almost all cities (green dots).

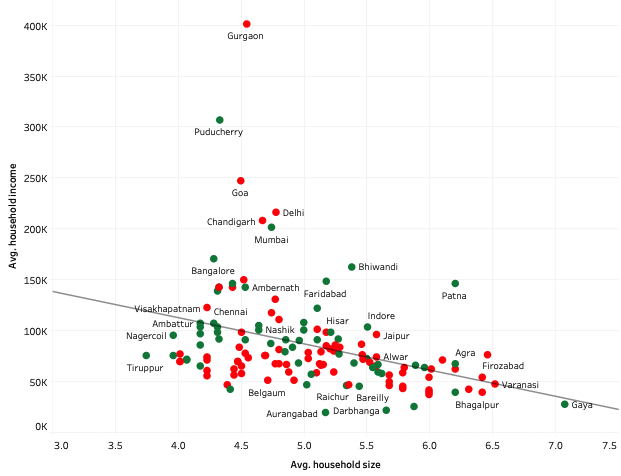
https://miro.medium.com/max/60/1*GGzEzsKq-1CSRjED_WCMJg.png?q=20



Avg. annual growth in employment in Agriculture, Industry and Services in India (2000–2011)

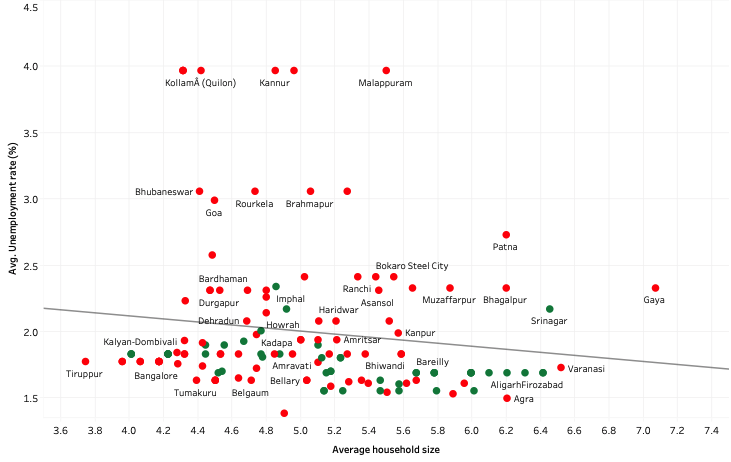
Average household size is negatively correlated with average household income. Increasing the size of the household by one member, decreases the mean income of the household by approx. Rs. 26000 a year, holding everything else constant. The red dots indicate a net increase in unemployment over the period in question, while the green dots indicate a net decrease in unemployment.

https://miro.medium.com/max/60/1*_3ghfhwjfZy9iGwputz_oQ.png?q=20



However, we find that as the average size of a household increases by one member, the unemployment rate decreases by 0.11%. This may point to the prevalence of family-run enterprises in India, where the entire family lives and works together.

https://miro.medium.com/max/60/1*vyrjK8KNBkCC7seX30rzCw.png?q=20



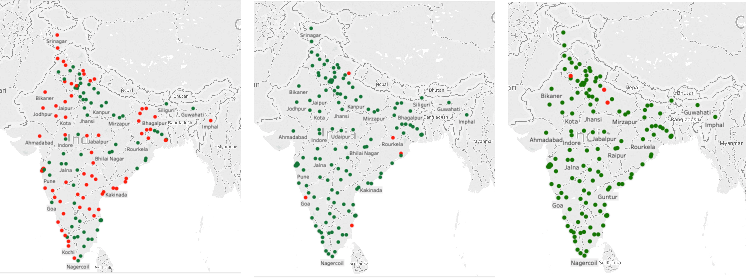
Taken together, the graphs may indicate that as the size of the household increases, members have a higher chance of finding employment, but the average income per person in the household actually decreases. Gurgaon, Lucknow, Raipur and Ghaziabad have recorded the highest average increase in unemployment between 2000–2011. Gurgaon witnessed an increase in annual unemployment despite the IT boom. Kholapur, Ratnagiri, Kannur and Mumbai has decreased unemployment between 2000–2011.

https://miro.medium.com/max/60/1*CHLbLJ9rqiIMk95tuROZ3g.png?q=20

**Model:**

Growth in employment in agriculture, industry and services reveals that most cities had a negative growth rate in the number of people employed in agriculture, while, employment in industry and services has grown in almost all cities (green dots).

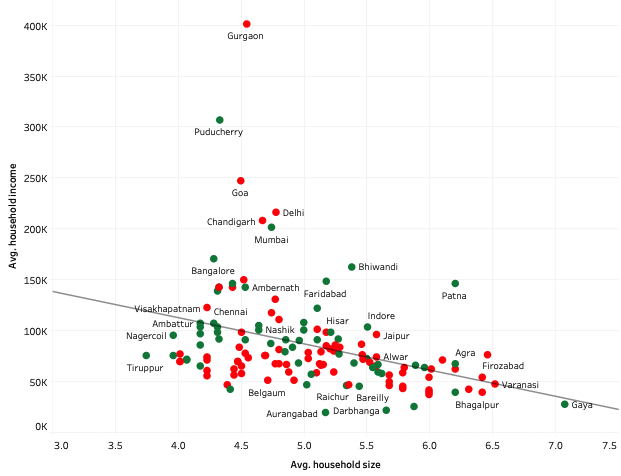
https://miro.medium.com/max/60/1*GGzEzsKq-1CSRjED_WCMJg.png?q=20



Avg. annual growth in employment in Agriculture, Industry and Services in India (2000–2011)

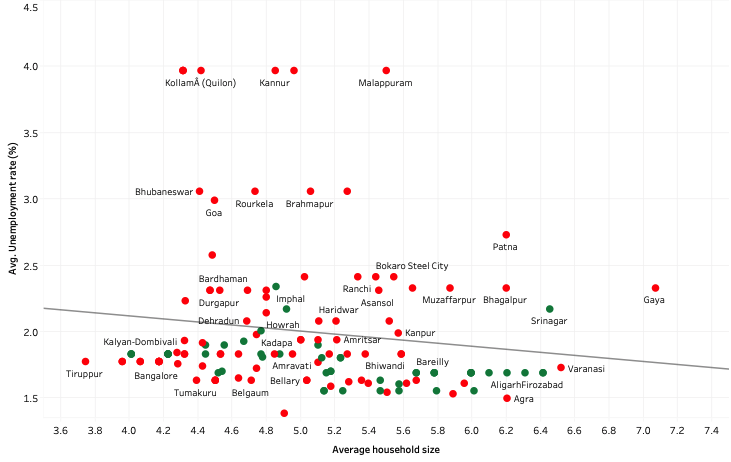
Average household size is negatively correlated with average household income. Increasing the size of the household by one member, decreases the mean income of the household by approx. Rs. 26000 a year, holding everything else constant. The red dots indicate a net increase in unemployment over the period in question, while the green dots indicate a net decrease in unemployment.

https://miro.medium.com/max/60/1*_3ghfhwjfZy9iGwputz_oQ.png?q=20



However, we find that as the average size of a household increases by one member, the unemployment rate decreases by 0.11%. This may point to the prevalence of family-run enterprises in India, where the entire family lives and works together.

https://miro.medium.com/max/60/1*vyrjK8KNBkCC7seX30rzCw.png?q=20



MODEL B: Factors that cause unemployment amongst youth

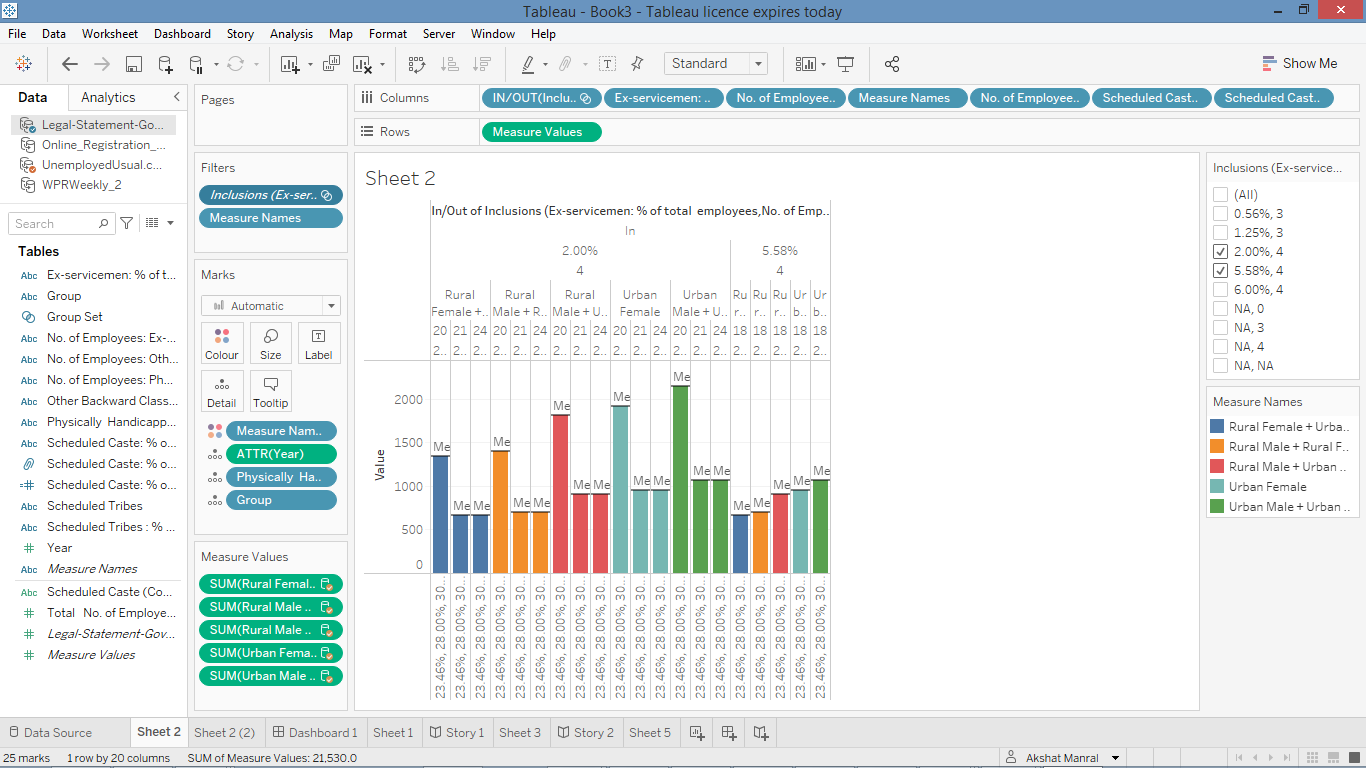
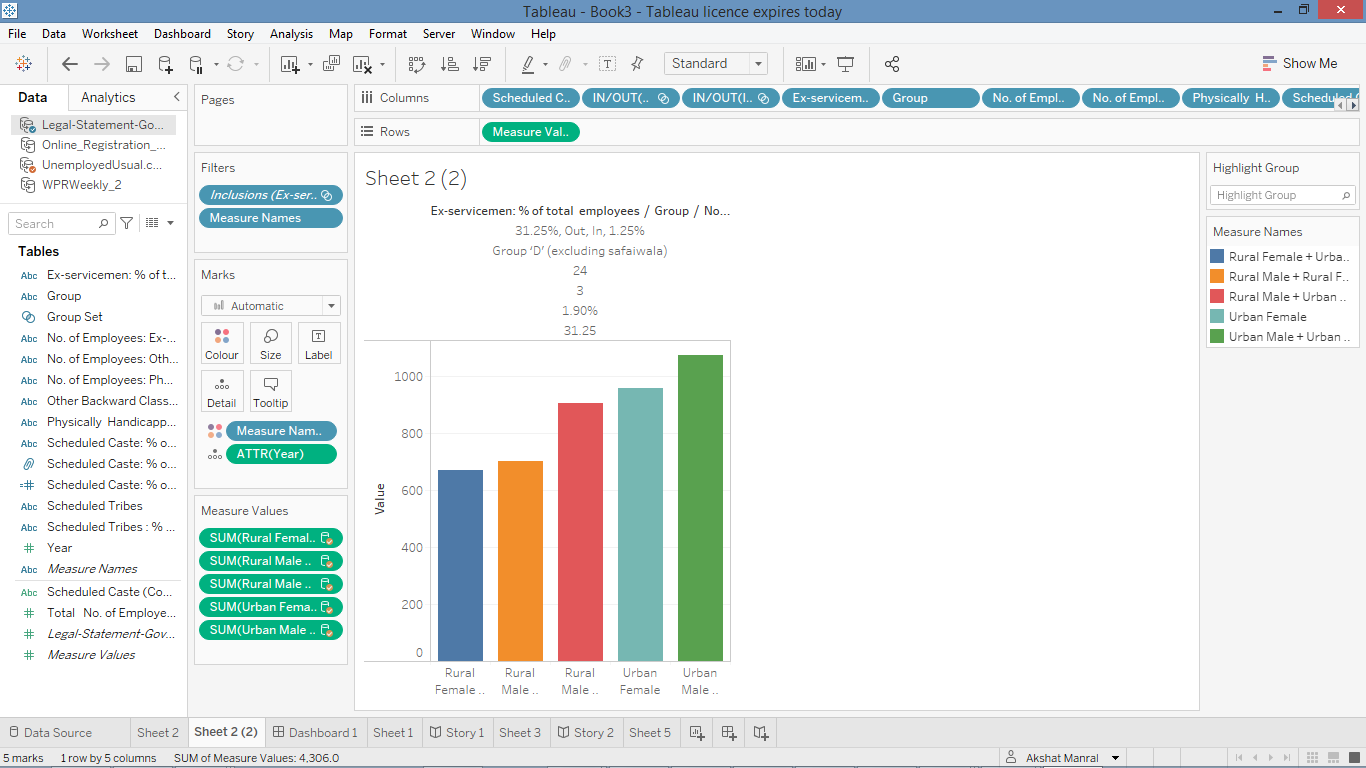
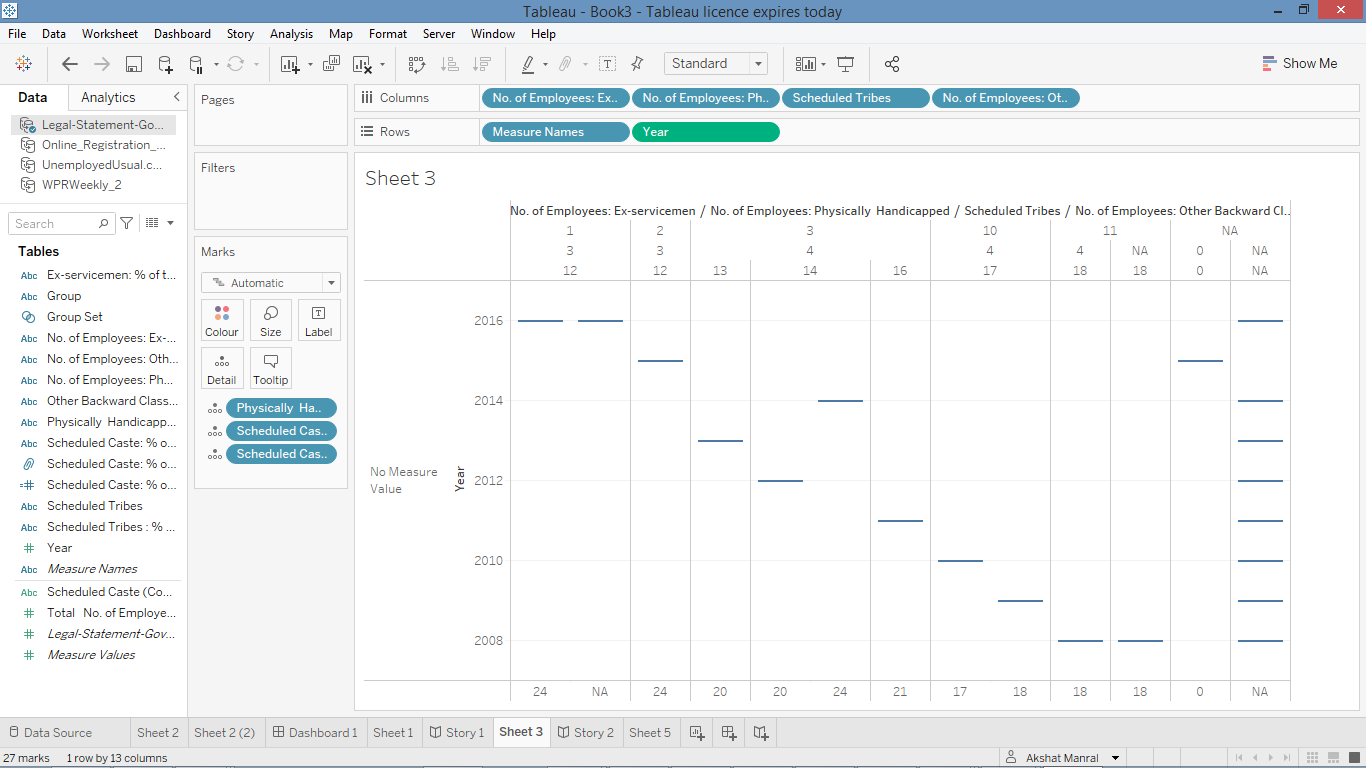
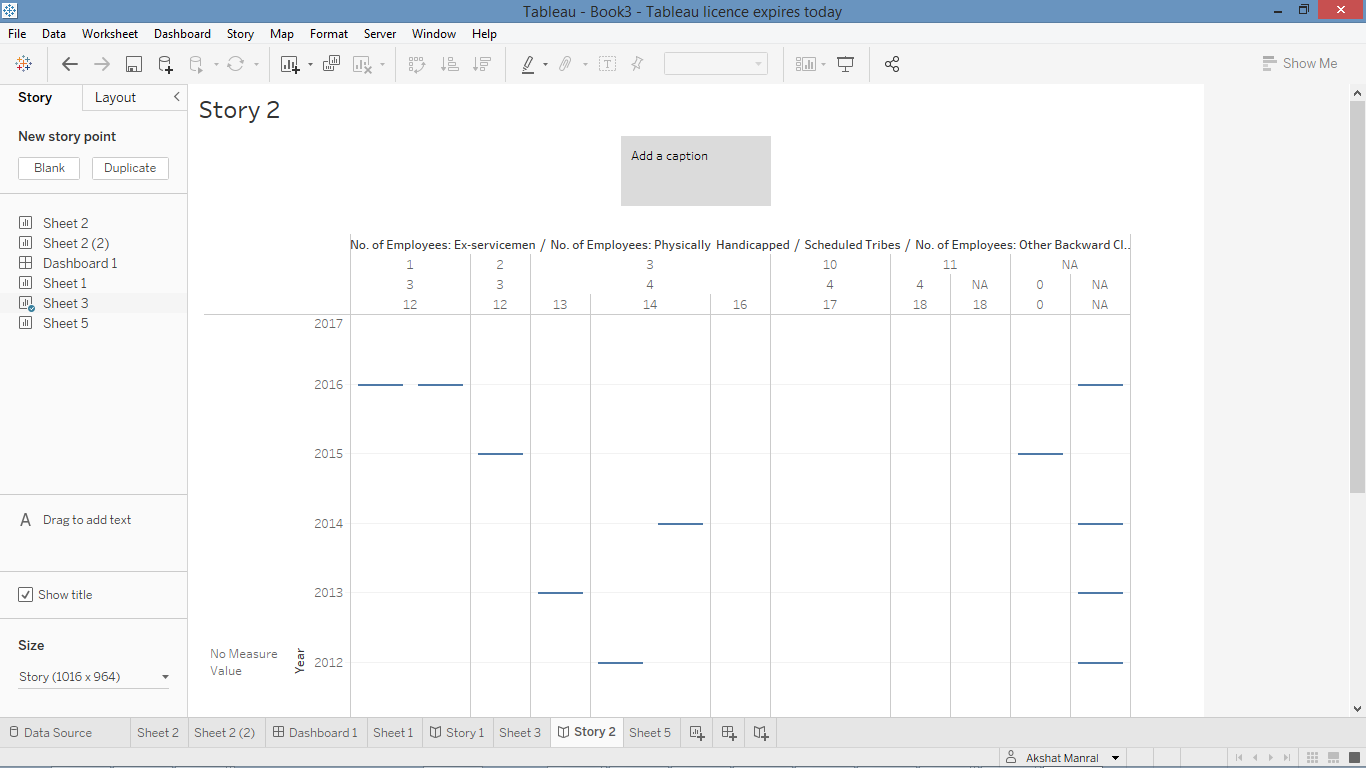
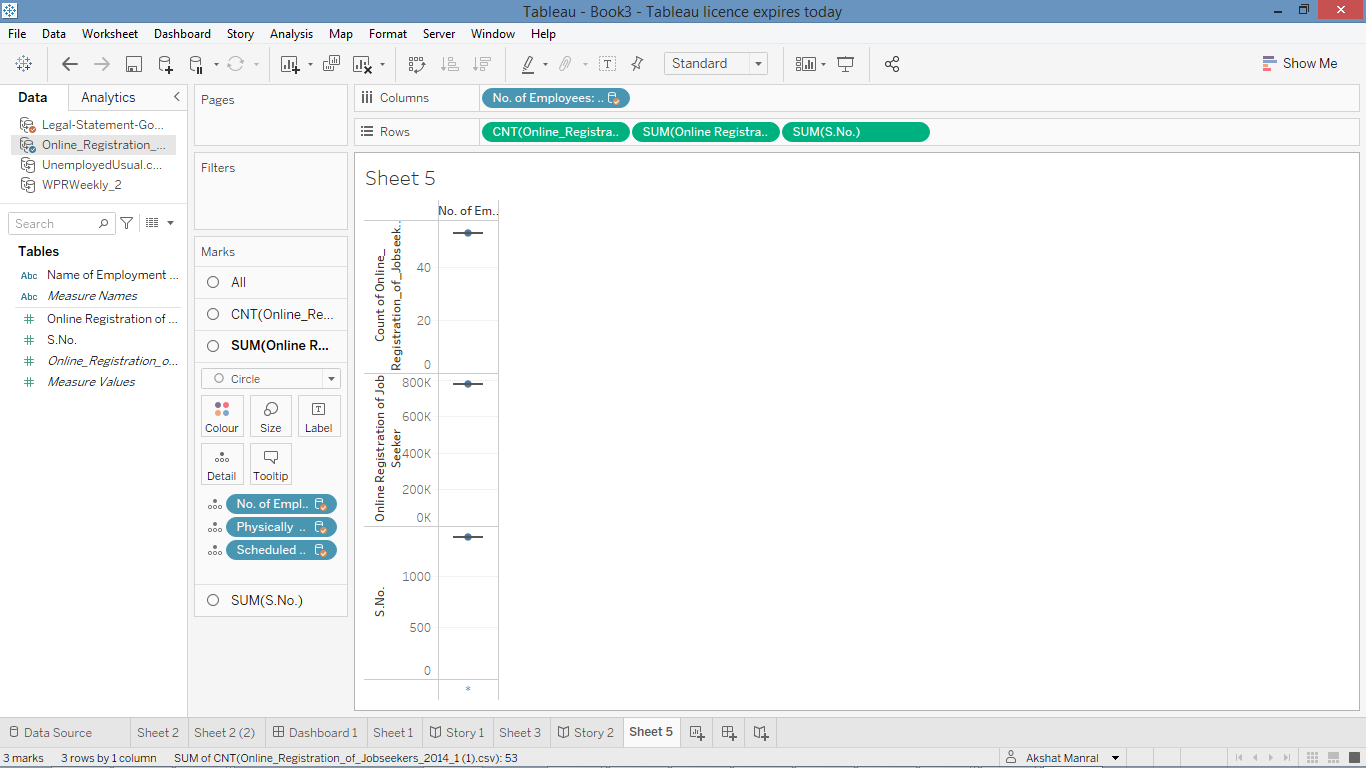
We defined two logistic regression models to determine the causes of unemployment:

Model 1: Unemployed ~ Age + Gender + Education + Social Group + Up skill + State + VT

Model 2: Unemployed ~ Age + Gender + Education + Social Group + Up skill + State + VT\_Field

The regression models were run on the original datasets (without over sampling), and diagnosed accordingly. Results are shown below:

**Results:**

****

**Conclusion:**

Unemployment among the youth is a cause of concern in India today. Rapid urbanization, as well as the slow rate of job growth has made unemployment an area that needs a pro-active intervention.

In this project, we studied the problem of youth unemployment in India using a literature review and exploratory data analysis in order to understand the domain, machine learning algorithms to determine whether unemployment is on the rise or decline in a city, whether a youth member is unemployed or not based on his/her demographics, followed by regression models to determine the causes of youth unemployment in India, as well as to evaluate the monetary impact of unemployment. This analysis was conducted first at a city-wide level (using MarketLine City Advantage dataset), and then at the individual level (using the 5th EUS Survey dataset).

Based on the findings, we arrived at a business plan for a social enterprise in the handicraft sector, which would rehabilitate disconnected youth and provide vocational training for selected youth in reviving the dying handicrafts of India. In so doing, we address 3 Sustainable Development Goals viz. no poverty, education and well-being and decent work and economic growth.

The models used in this project can be reused when new data becomes available from the Labour Bureau and MarketLine City Advantage. The data will need to be cleaned and preprocessed before running the models on them. Also, if the survey questionnaire is updated, then the models will need to be adjusted accordingly.

# References

Azad India Foundation. (2019, Jan 1). BEGGARY IN INDIA. Retrieved April 7, 2019, from Youth Ki Awaaz: <https://www.youthkiawaaz.com/2008/04/beggary-in-india/>

Belen V.M., C. A. (2013/14). Statistical Analysis of Unemployment in Europe. Technische Hochschule Nürnberg Georg Simon Ohm, Nürnberg.

Belen Villena Maria, C. A. (2013/14). Statistical Analysis of Unemployment in Europe. Technische Hochschule Nürnberg Georg Simon Ohm, Nürnberg.

BI India Bureau. (2019, Mar 26). Job creation in India slows down by 6.9% in January: ESIC payroll data. Retrieved April 22, 2019, from Business India Insider: <https://www.businessinsider.in/job-creation-in-india-slows-down-by-6-9-in-january-esic-payroll-data/articleshow/68574896.cms>

BI India Bureau. (2019, Feb 6). Unemployment in India is so bad that over 150 MBAs and engineering graduates applied for a handful of sanitation jobs. Retrieved April 22, 2019, from Business India Insider

Central Statistical Organization. (2008). National Industrial Classification (all economic activities). Ministry of Statistics and Programme Implementation. New Delhi: Government of India.

Development Commissioner (Handicrafts). (2019, March 20). Schemes. (M. o. Textiles, Producer, & Government of India) Retrieved April 7, 2019, from Development Commissioner Handicrafts: <http://www.handicrafts.nic.in/>

Economic Times . (2014, Nov 18). India has world’s largest youth population: UN report. (Economic Times, India Times) Retrieved April 26, 2019, from Economic Times: <https://economictimes.indiatimes.com/news/politics-and-nation/india-has-worlds-largest-youth-population-un-report/articleshow/45190294.cms>

Frey C. B. & Osborne, M. A. (Sept 2013). THE FUTURE OF EMPLOYMENT: HOW SUSCEPTIBLE ARE JOBS TO COMPUTERISATION? [https://www.oxfordmartin.ox.ac.uk.](https://www.oxfordmartin.ox.ac.uk./)

Handicrafts.NIC. (2019, Jan 1). List of Identified as Endangered Craft. Retrieved April 2019, from <http://handicrafts.nic.in/pdf/List_Of_Identified_As_Endangered_Craft.pdf>

HULT Prize. (2018). 10th ANNIVERSARY CHALLENGE: FOR US, BY US YOUTH UNEMPLOYMENT. Retrieved April 23, 2019, from HULT Prize: <http://www.hultprize.org/challenge/>

India CSR Network. (2013, May 22). The Declining Legacy of India — Rural Artisans: Report. Retrieved April 07, 2019, from IndiaCSR: <https://indiacsr.in/the-declining-legacy-of-india-rural-artisans-report/>

Kassambara. (2018, March 11). Multicollinearity Essentials and VIF in R. Retrieved April 21, 2019, from STHDA — Statistical tools for high-throughput data analysis: <http://www.sthda.com/english/articles/39-regression-model-diagnostics/160-multicollinearity-essentials-and-vif-in-r/>

Kedia, S. &. (2018, Oct 18). 5 ways tech can transform the future of work in India. Retrieved April 23, 2019, from World Economic Forum: <https://www.weforum.org/agenda/2018/10/five-ways-tech-can-transform-the-future-of-work-in-india/>

KPMG. (2016, Jan 1). KPMG Environmental Scan 2016. India. Retrieved from <http://www.tsscindia.com/media/2565/kpmg-environmental-scan-report-2016.pdf>

Labour Bureau. (2019, January). 5th Employment Unemployment Survey Data. Chandigarh, India.

Labour Bureau. (Sept 2016). Report on Fifth Annual Employment — Unemployment Survey (2015–16). Ministry of Labour and Employment. Chandigarh: Government of India.

Labour Bureau. (Sept 2016). Report on Youth Unemployment — Employment Scenario Vol 2. Ministry of Labour & Employment, 5th Annual Employment Unemployment survey 2015–16. Chandigarh: Government of India.

MarketLine Advantage. (2017, June 1). City Data. MarketLine Databases . London, United Kingdom.

Mehta, P. (2019, Jan 1). Main Causes of Unemployment in India. Retrieved Feb 14, 2019, from Economics Discussion: <http://www.economicsdiscussion.net/articles/main-causes-of-unemployment-in-india/2281>

Mehta, P. (2019, Jan 1). Main Causes of Unemployment in India. Retrieved Feb 14, 2019, from Economics Discussion: <http://www.economicsdiscussion.net/articles/main-causes-of-unemployment-in-india/2281>

Ministry of Labour and Employment. (2019, March). Round 5 — Schedule B. Annual Employment-Unemployment Survey 2015 . Government of India.

Minstry of Labour & Employment,. (2004). ALPHABETICAL INDEX OF N.C.O., 2004 WITH OCCUPATIONAL TITLES AND EQUIVALENT CODE NUMBERS OF N.C.O., 1968. Government of India.