Final Project Report

Name: Srimant Mishra
Data of Submission: 9th December 2025

Introduction / Motivation of this project

India, for a majority of the period since its independence, has been a cash based economy. High tax rates, large compliance burden, and an informalized structure of the economy perpetuated this trend for several decades.

However, in 2016, there was a major change in the tax code of India. The “Goods and Services Tax” got implemented in 2016 and this led to a whole series of changes that started to transform the nature of the formal economy.

For starters, the GST subsumed 17 major taxes and a number of cesses. This directly reduced the compliance burden for entities engaged in business. Then, it started providing incentives in the form of input tax credit (for the production of intermediate goods). This made it financially lucrative for firms to start doing business in the formal manner and avoid cash-based transactions. In fact, large businesses started using GST registration and compliance as a precondition to deal with small businesses - creating a type of business pressure which led to an acceleration in the formalization of the economy.

Moreover, along with the broad suite of changes that the GST created (increasing formalization and registration with the tax authorities), a novel development was also the E-Way bill. Put simply, this is a document that contains details about the parties involved in trade, the items that are being moved, the value of the goods traded, and the mode of transport being utilized. Each E-Way bill has a 12 digit unique identifier and is able to capture shipments for sale and non-sale reasons, as well as exports and imports.

These E-way bills, generated from 2017, have provided a new set of eyes to analyze the Indian economy. They provide insight into how goods move from state to state, where activity has been concentrated, which states have adopted the system faster than others, etc. Exploring these trends and linking them with employment metrics in India is an exploration that I was interested in performing.

Besides the above point, the relatively novel nature of this project was an added reason for me to pursue this project. While other individuals and agencies have done explorations with this data over the past few years, in the grand scheme of things, due to its recency, complexity, and niche nature, not much exploration of this information has taken place outside of the government or journalistic sector. And even in those cases, the insights that the government sought to obtain from the data are divergent from my own.

Background / Related Work

There has been a lot of governmental and journalistic work on the E-way bills in the previous years. Some of these and their major conclusions are listed below.

Report: India on the Move

Relevant Findings:

- This report talks about the introduction of the E-way bill system and the platform on which the backend works.
- It asserts that E-way bills provide data that is useful for tracking both intra-state and inter-state supplies (which helps track the volume of trade over time).
- Using E-Way bills a Net Trade Balance for each state can be created, which is useful for various policy discussions.
- 58% of E-Way bills (2018 numbers) are for goods travelling within 200 km, suggesting that most trade happens locally (gravity effect).
- 17% of bills cover distances over 1000 km.
- The report examines the average number and value of E-way bills per supplier to understand the spread of industrialization and formalization within a state's economy.

Report: Economic Survey 2022-2023

Relevant Findings:

- The E-way bill system along with the Goods and Services Tax Network (GSTN) has been instrumental in enabling the formalization of business transactions.
- E-way bills are a technology backed tax governance reform that has enhanced tax compliance and improved fraud detection systems.
- E-way bills (as a technology backed reform) has led to revenue buoyancy due to the ease in collection and finding fraud using AI.

- The E-way bill network has generated trust for small businesses through monitoring compliance.

- The E-way bill network is part of the overall suite of Digital Public Infrastructure that the government is pushing for. Taken together, it will help create greater value for money for consumers, while reducing the compliance burden for producers.

Investigation: Fake Invoices are being generated

Relevant Takeaways:

- While the system is seeing greater adoption, there are also ways to fool the system. Thus, all the bills that are being generated should not be considered genuine. In many cases there is a round-tripping of goods to get input tax credit.

Report: Assessment of Logistics Cost in India

Relevant Takeaways:

- The government itself uses E-way bill data to identify key trade routes and origin-destination pairs across the country.

- The government tracks movements using E-way bills in 9 categories (from below 200km to over 2500km).

- E-way bill data allows the creation of regional zones for granular spatial analysis.

- There is a high level of inward and outward movement in the western region of India due to lower logistics cost and “backhaul availability” (trucks don't return empty handed).

- While the compliance levels for Road have been good, compliance for other modes of transport like Rail, Air and Water have remained weak.

Datasets used

E-Way bill data

1. Contains details of merchandise trade above a certain threshold (Rs. 50000).
2. Segregates data by year, month, and state.
3. Splits the trade data into 3 main categories - Outgoing, Incoming, and Within-State trade.
4. Contains details for each state going back to 2018 (9500+ rows)
5. Lists the number of bills generated by a state in a particular time period
6. Also provides the value of the E-way bill generated (aggregate) for a particular state in a particular month.

Employment Data

1. Contains data for the time period 2017-2023.
2. Provides employment status for men, women separately, as well as for "all persons" combined.
3. Provides granular data for all states and union territories (14000+ rows)
4. Breaks down employment categories into 5 categories - "All Self Employed", "Casual Labour", "Regular Wage / Salary", "Self-employed helper in household", "Self employed own account worker"
    - Within these categories, my analysis is centered around Casual labour and Regular wage / salary employment.
5. Employment data is also segregated into Rural, Urban, and Rural + Urban. I have relied on the Rural + Urban data to do my analysis.

The datasets listed above are made available under the Government Open Data License (https://www.data.gov.in/Godl) which gives royalty free access to use, adapt and publish the data provided.

Explicit Research questions / hypotheses

While I planned to do a variety of exploratory data analysis and create visualizations to see what insights could be gleaned from the data, the following were my explicit research questions:

1. Is there a geographical North-South or East-West divide in India when it comes to merchandise trade?

    - The answer to this would help further the discussion that is taking place in many other forums about a geographical split in India's overall metrics (South being more literate than the north, North having more fertility than the South etc.)

2. Is there increasing formalization in India's merchandise trade economy over the years (proxied by a growing usage of E-way bills)

3. What is the level of growth of merchandise trade (as discerned by the E-way bill data) across India? Are there regions which are growing faster in their usage of this system than others?

4. What are the correlations (if any) between merchandise trade levels, and casual or formal employment.

5. What is the link between merchandise trade and employment status (casual and formal) for men and women.

Finally, I also wanted to understand a bit more about each state and for this I aimed to create a state-level dashboard, where the key metrics for any state could be seen and contextualized (through comparisons with national statistics). I also aimed to provide some useful information at a glance about the state overall in this dashboard (at a high level).

Methods used

To answer my research questions and build the dashboard that I intended to, there were several things that I did. In no particular order, these include:

1. Extracting state level data (from the E-way bill dataset), and storing them in one data structure (dictionary) for easy reference. The information that I stored in this data structure included directly extracted metrics (Number and Value of E-way bills in a particular year or month for a particular state), as well as derived metrics such as correlations (between incoming and outgoing trade, between internal and external trade, between value and count of E-way bills, etc.)

2. Developing time series charts to visually see what the trend lines for various trade and employment metrics are, and how they evolve over time.

3. Calculate various CAGR statistics (Compounded Annual Growth Rate) to see how the multi-year growth trend for various metrics has been (count or value of incoming, outgoing, and internal trade e-way bills, among others)

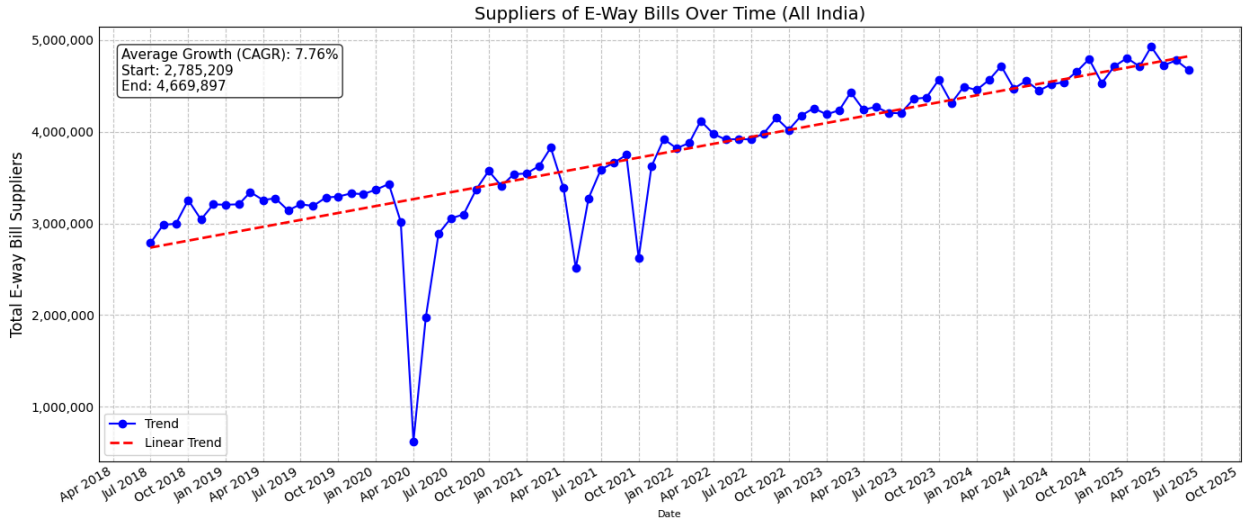
4. Using Geopandas to overlay the data that I generate onto maps. For this I relied on an external GeoJSON data source which contained the latest mapping information for India. It contains information for not just states but also districts in India.

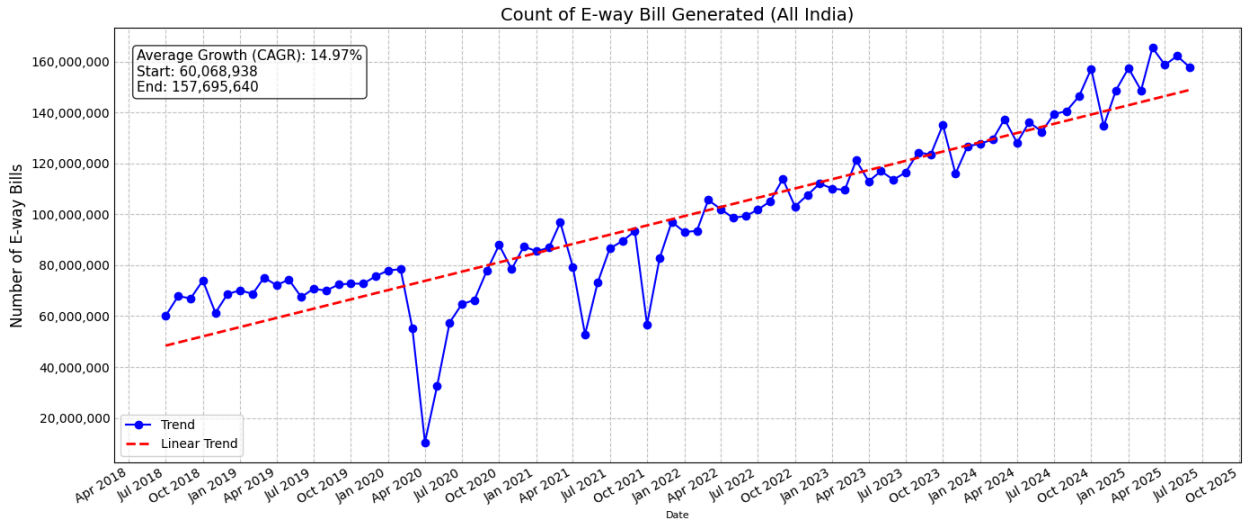
    - Superimposition of the data onto the maps that I generate using this GeoJSON data will be crucial in answering my research question about East-West or North-South divides when it comes to various metrics.

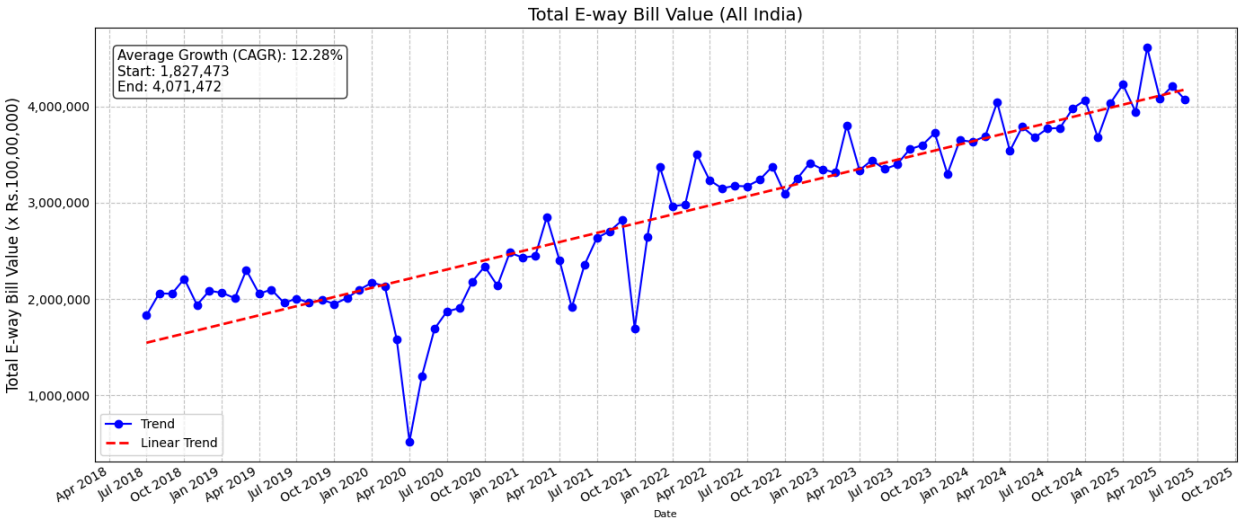
Summarization of findings

The results shared below answer the research questions presented above, but go in a slightly different order than the order of the questions above.

Key Finding #1: There has been high and sustained growth in merchandise trade over the previous 6 years (the range that we can discern from the data).







- The above graphs show that the number of suppliers (first graph), the quantity of e-way bills generated (middle graph), and the value of the e-way bills generated (final graph) have all grown at a very high rate.

- The CAGR for each of these (supplier CAGR at 7.76%, count growth at ~15%, and value of e-way bill CAGR at 12.28%) have all outpaced the overall rate of growth in India (which was around 5.5% for this time frame).

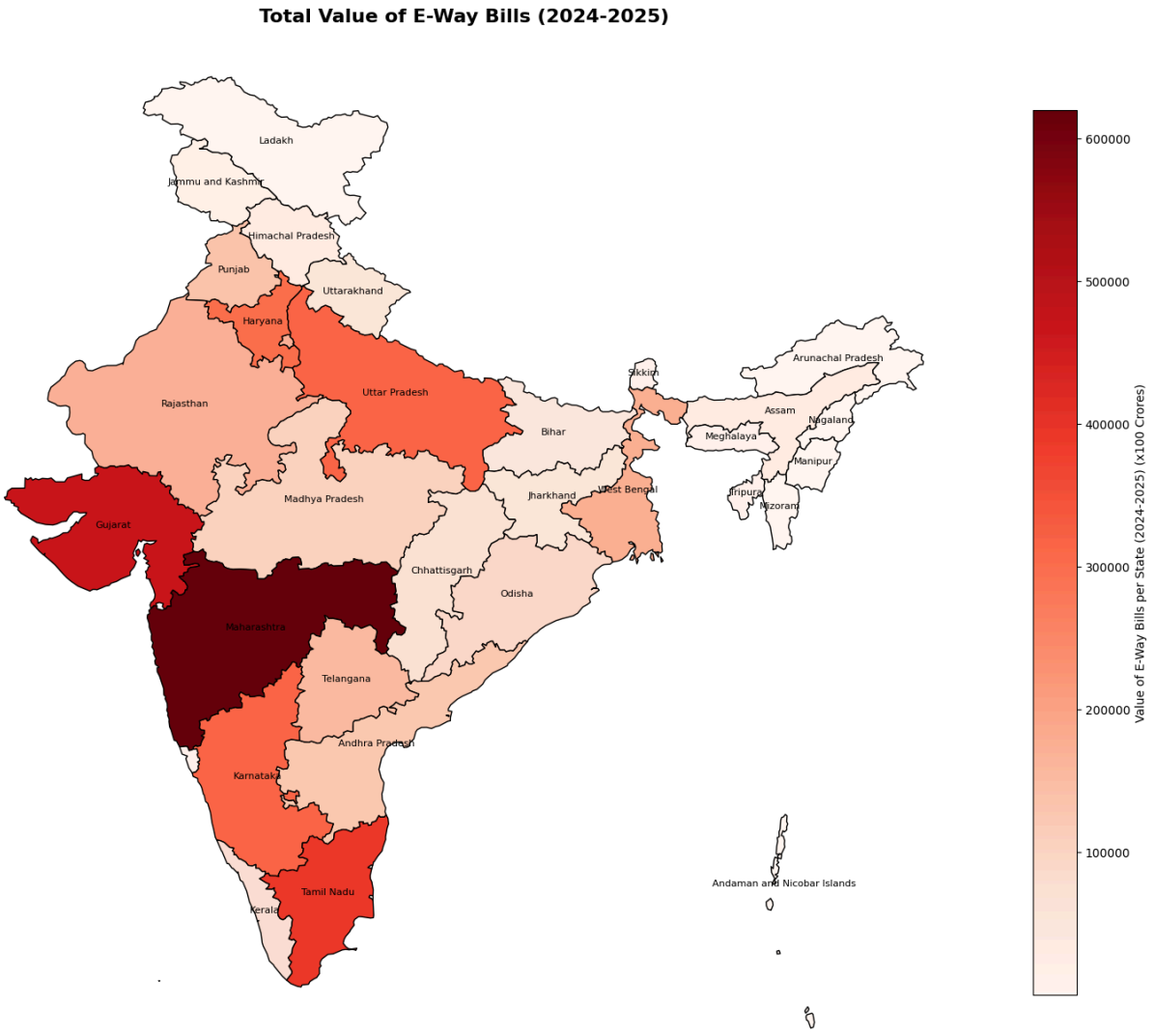
- The count of e-way bills has grown faster than the value growth. This indicates that businesses are using this system more over time for smaller transactions (indicating comfort with the system).

- The very high levels of CAGR in the above charts are notable for the fact that this period saw the Covid-19 pandemic as well, which was a major shock lasting over 18 months where the growth was negative. Thus, it is very interesting to see the momentum in the bulk merchandise trade overall in this period.

- Indirect / Derivative conclusions that we can generate from these graphs include:

    - There must be a robust manufacturing, and export-import growth in the country to fuel the rise that we see.
    - More businesses are choosing to shift into the formal system of movement of goods, which is a break from decades old attitudes which drive businesses to work informally to avoid tax.

Key Finding #2: There is a clear West-East divide when it comes to the generation of e-way bills (taken as a proxy for bulk merchandise trade). However, the divide is not very neat and there are some notable exceptions to this pattern.



- The state of Maharashtra clearly stands up to its reputation as the economic capital of India. It generates the highest value of e-way bills (aggregate) in the country as a whole. This is partially explained by its large GSDP, its role as an export-import hub, and a rich consuming class of people there.

- Gujarat is the second most active state when it comes to generating bulk merchandise trade, also owing to its important role as an export-import hub, and the presence of large ports.

- The state of Tamil Nadu in the far south comes third in terms of the value of E-way bills generated. This is better contextualized by keeping in mind the fact that it has been one of India's fastest growing states in the recent past, and is a leader in the country when it comes to the manufacture of heavy machinery and textiles.

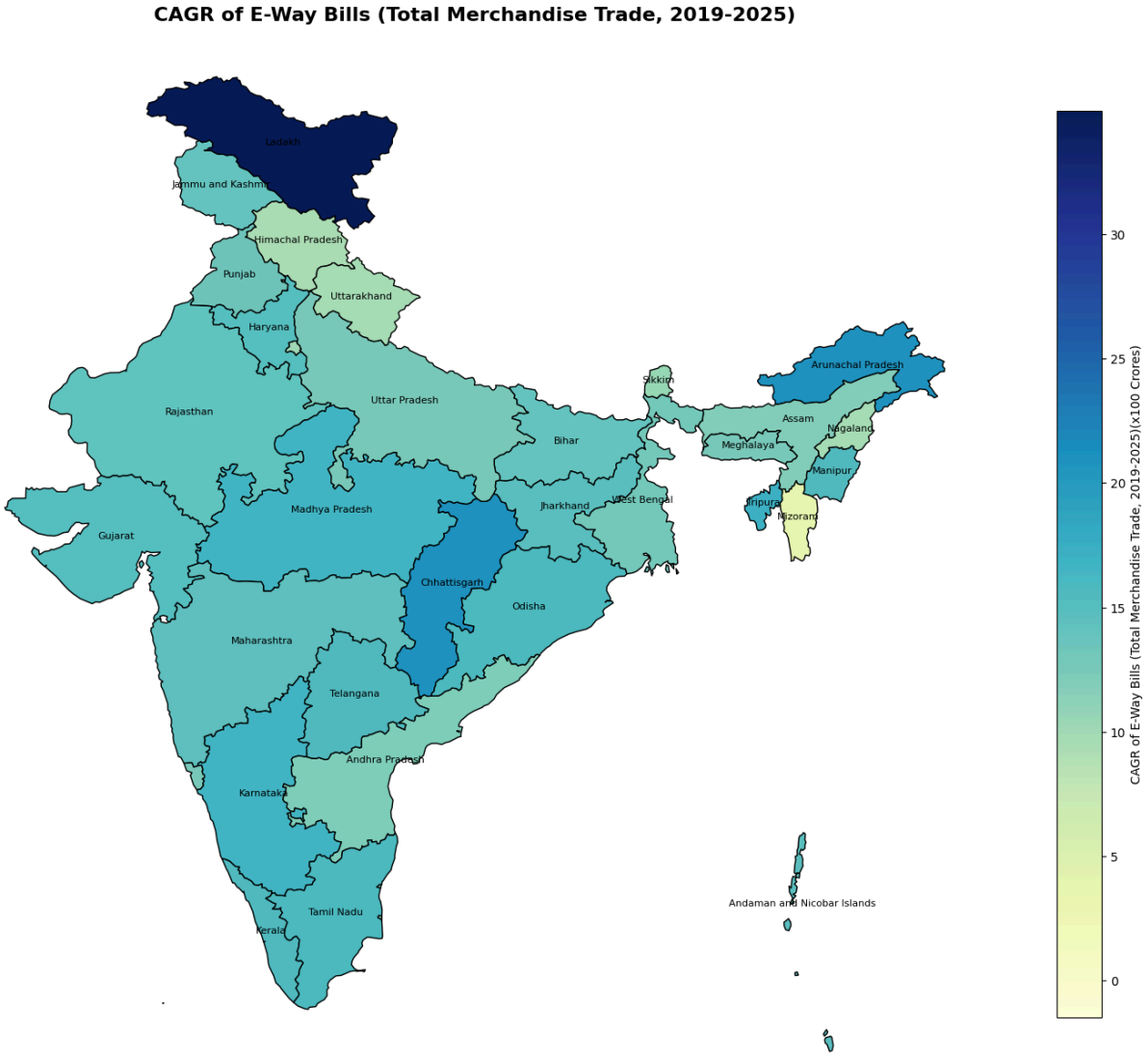
- The West dominates in its generation of e-way bills overall. However, the state of Kerala (far south besides Tamil Nadu) is an interesting exception. The state has been known for its high social indicators for decades now (healthcare, education, metrics determining nutrition and maternal outcomes for women, etc.). Despite this, it has not been able to translate its positive social outcomes into economic outcomes.

- In the east there is broadly much lower bulk merchandise trade happening. However, despite this, the state of West Bengal does see some strong e-way bill generation happening. More research is needed into this field to explore how much of this is due to new Capex investment that has taken place over the previous decades, and how much is owed to the historical legacy of the state as one of the great trading regions of the world (in the mid 20th century).

- The North-western and northern (Himalayan) states are major laggards. Poor infrastructure and lower populations could explain some of this. However, more research is also needed to see if there are policy bottlenecks that are inhibiting the rise of these states in terms of bulk goods trade.

- The mineral belt of India (Odisha + Bihar + Chhattisgarh + Jharkhand + Madhya Pradesh) see surprisingly low merchandise trade occurring. This is a counterintuitive finding given the mining activity that takes place there (which should ideally be generating very high value e-way bills).

Key Finding #3: There is growth occurring all over the country in merchandise trade. Some regions are rising faster than others but there is no region experiencing a decline (there is stagnation in one state of the North-east, however)



- Most of the states and union territories have seen double digit growth in India (with a few exceptions like the Himalayan states of Himachal Pradesh and Uttarakhand).

- Mizoram, in the north-east, interestingly, has seen a stagnation in its bulk merchandise trade. This is something that needs to be studied further, as it has not experienced any significant shocks that others in the region have seen (other states have seen internal conflict, military law in Manipur and Nagaland, spillover effects of the conflict in Myanmar, etc., and still managed to grow). Therefore, it is

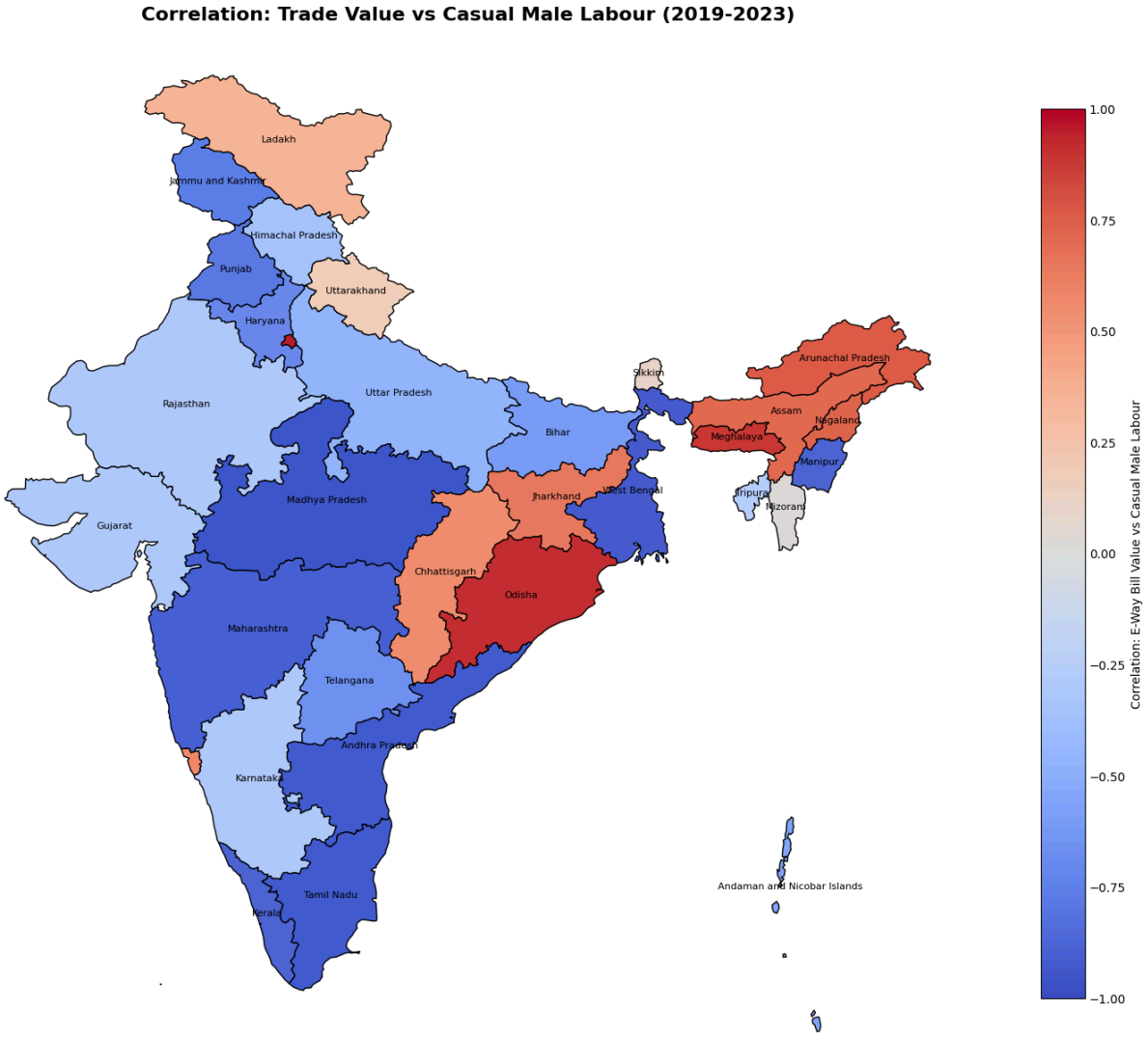
unclear what is the root cause behind the stagnation in Mizoram. One possible explanation could be that the infrastructure push in the northeastern region has reached other states but not Mizoram in a meaningful way so far.

- Ladakh in the far North has seen the fastest growth in its merchandise trading activity. This, most likely, is powered by various things such as a low base + push to create infrastructure there for military purposes + autonomy due to the J&K Reorganization Act, 2019 + rising connectivity links with other areas in the region. Despite this, how it managed to reach over 30% CAGR is something that should be studied more.

- Chhattisgarh and Arunachal Pradesh are also 2 interesting finds. Their higher growth in bulk merchandise trade could be linked to things like successful counter insurgency operations in Chhattisgarh, creating an environment that is amenable to more business activity. Arunachal Pradesh, the Himalayan state in the north-east sharing a border with China, has also seen a lot of infrastructure creation due to military exigencies and started out with a low base - possibly explaining why it has been growing faster than the rest of the country.

Key Finding #4: A rise in merchandise trade generally leads to a fall in casual employment for men and women. However, there are notable exceptions to this.

The trend for Men



- Casual work can be understood as work with no fixed employment status (you are called to work based on the availability of work), no benefits, and no paid leave. For most of the country, we can see that there is a negative correlation between casual work and an increase in bulk merchandise trade (more e-way bill generation).

- However, for states in the East (Odisha + Chhattisgarh + Jharkhand), the Himalayan states like Ladakh and Uttarakhand, and for rich regions like Goa and the National Capital Region of Delhi, there is an inverse correlation. This implies that when more bulk merchandise trade takes place there, there is more casual work generated for men (instead of more formal, salaried work).

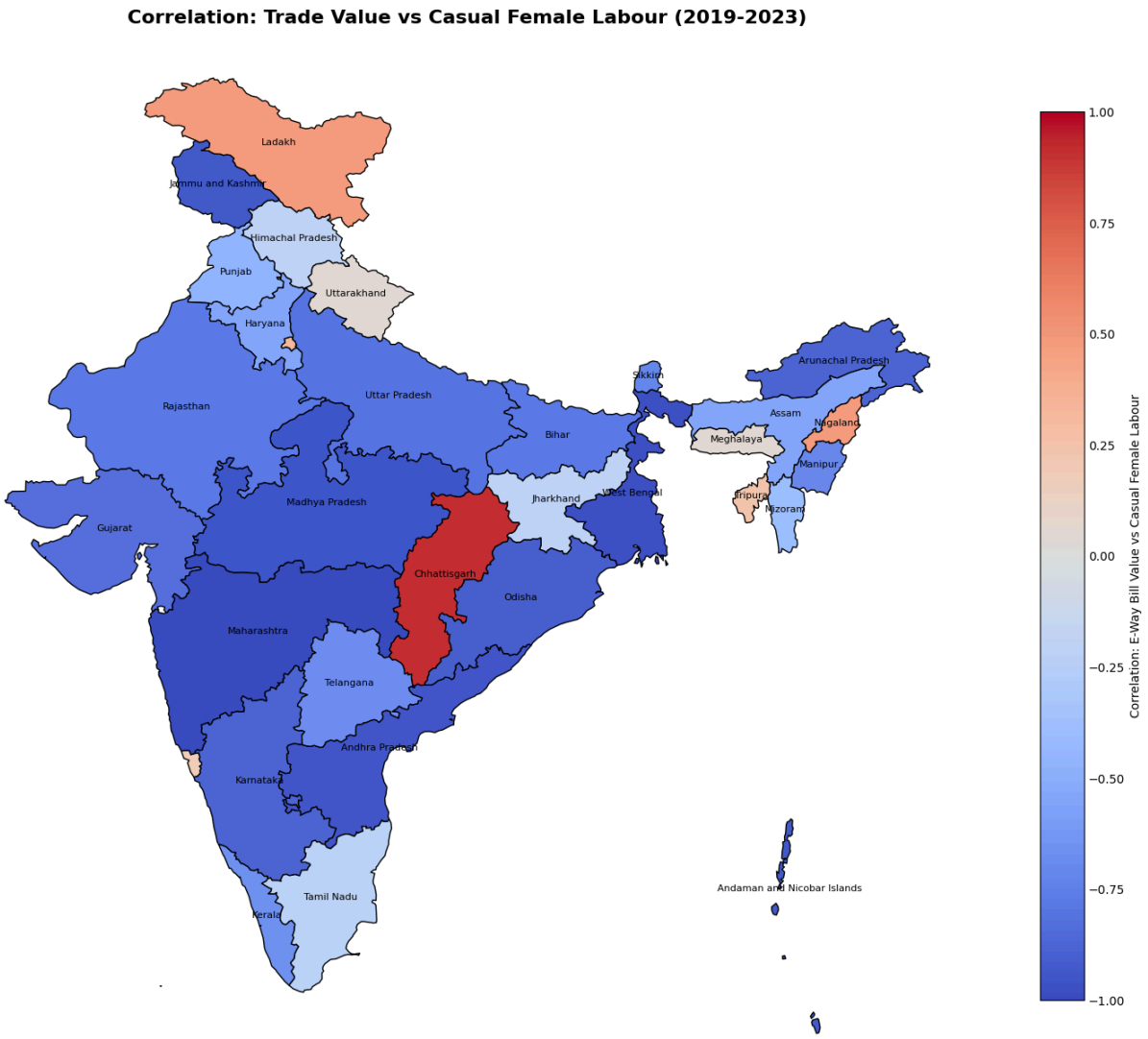
- The reason why eastern India could be seeing this trend (in the mining states of Odisha, Jharkhand, and Chhattisgarh) is that the local economy there is not set up in a way that allows easy creation of formal employment. Mining produces demand for manual, casual labour for excavation, loading and transportation in an ad hoc fashion in these states. Moreover, the mining companies could also be used to this pattern of employment and prefer this - so as to save on costs.

    - Another factor that could be driving the rise of casual labour is the fact that these are labour exporting states. The more skilled people from these states leave and go to other states within the country and abroad for better economic prospects. It is the remaining lower skilled male labourforce that is generally absorbed into the work that rises due to rising merchandise trade, thus leading to more casual labour (often via contractors).

- The northeastern states have the same migration story as the eastern mining states. However, those states are experiencing a much faster rise of construction over the past several years due to various government initiatives. This type of construction driven employment also attracts informal or casual employment in such areas.

- Finally in the richer regions like Delhi and Goa, more bulk merchandise activity being correlated with more casual labour might be due to the fact that those regions already have a very high level of formal employment. Since people who are skilled are already engaged in formal jobs there, more merchandise trading activity is generally performed by migrants coming to those regions - accepting any form of job that is available, out of desperation.

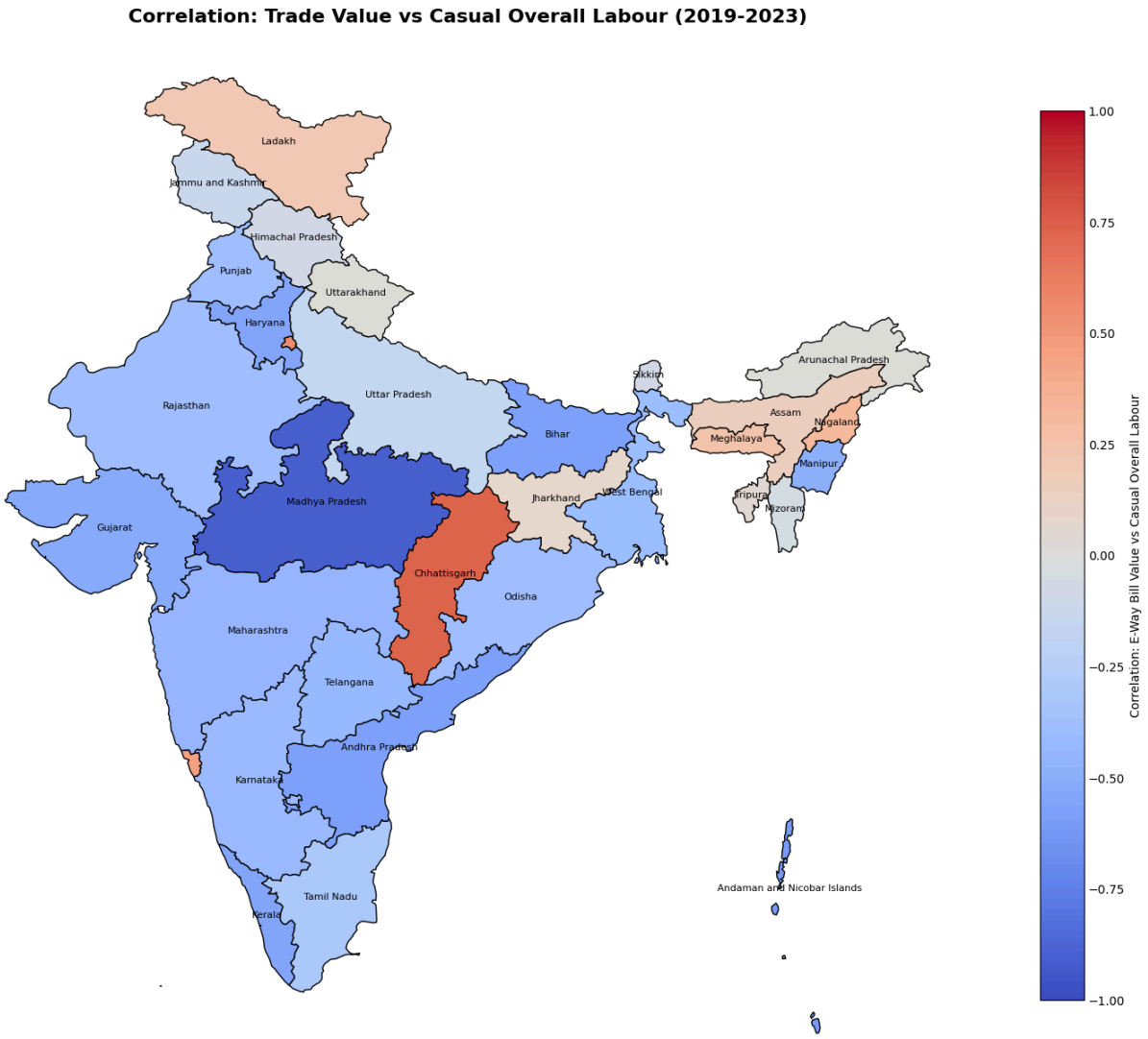
The trend for Women



- The trend for women is much more straightforward compared to men. Almost all the regions in the country see less casual employment for women when there is a rise in bulk merchandise trade.
    - As most women in India, who are not in salaried roles and seeking employment, remain engaged in agriculture or cooperatives led economic activities, there are very few women across the country who are seeking casual work in industries that come downstream from bulk merchandise trade.

- The clear glaring exception to this trend is the state of Chhattisgarh. This could be due to various reasons.
    - The first among these is that the state is dominated by Scheduled Tribes who lived along the edges of dense forests. Their livelihood is generally based on Minor Forest Produce (like gathering Tendu leaves), or working in mining. Due to this, more merchandise trade could be related to women doing casual work, as this type of work is the main source of livelihood for many families.
    - The second reason could be the U-shaped curve of female labour supply. This theory asserts that when there is a lot of distress, women go out and seek casual employment to support the family. However, as soon as there is some level of stability in the household, women withdraw from the workforce to take care of things like childcare and domestic duties. Chhattisgarh, which has been economically backward for decades owing to significant internal conflict (Naxalism and Maoism) could be experiencing such distress, leading to more women to seek employment there.
    - The third is cultural norms. In tribal communities, unlike many others, there is a general acceptance of the fact that women will work alongside men. This cultural practice could also be part of the reason why women, alongside men, sought casual employment.
- There are also a few other weekly correlated states or neutral correlation states for this trend including some in the north-east, Uttarakhand and Ladakh, Delhi NCR and Goa.

The trend Overall (All-Persons)

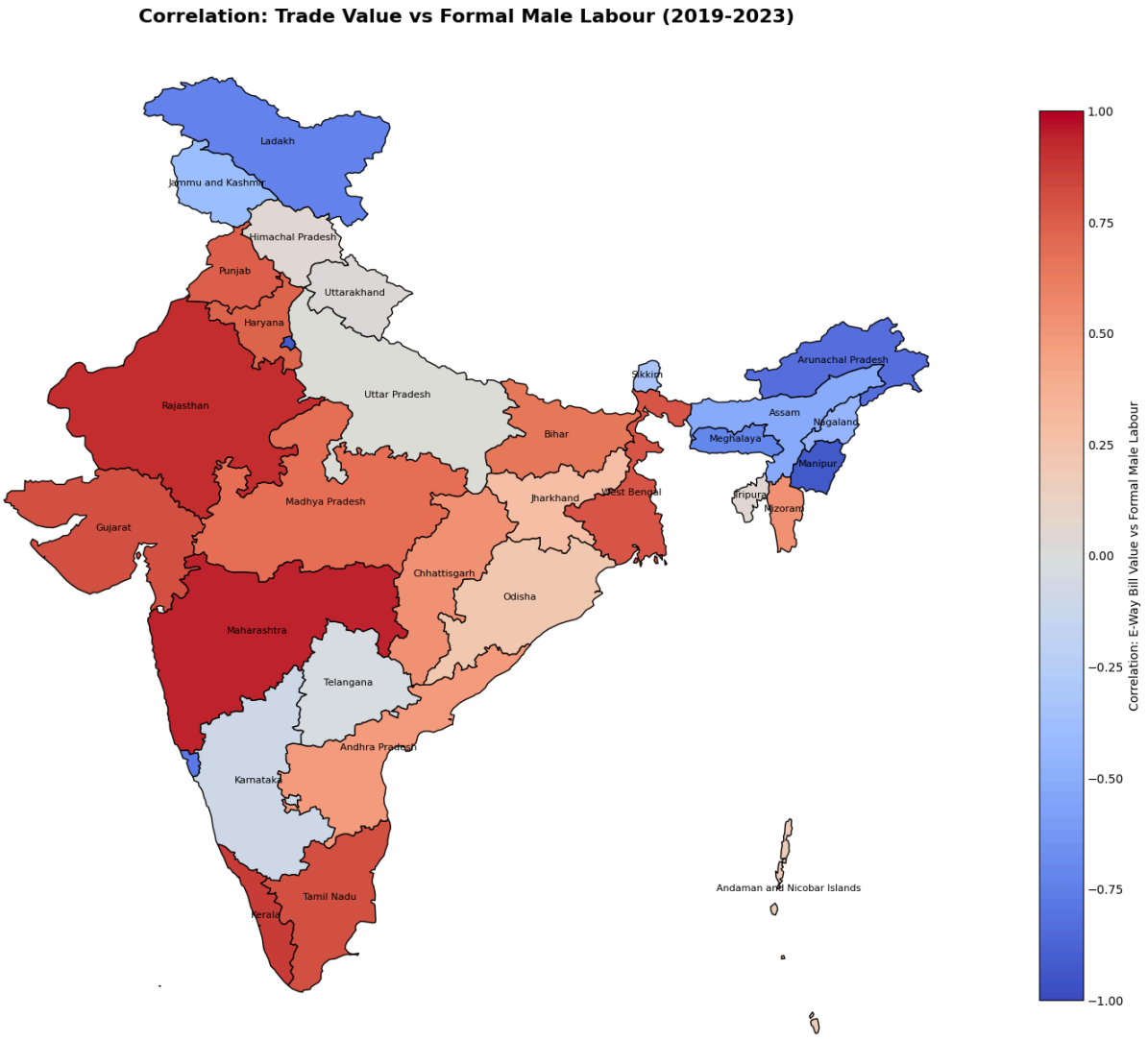


- Overall, we see that the country either has a strong correlation to decreasing casual labour as trade increases, or that there are some pockets where the correlation is neutral (Jharkhand, North-eastern states, Ladakh).

- The only regions that see a positive correlation (more casual labour coming from more merchandise trade) are Delhi & Goa (possibly due to near full employment for the eligible work-seeking population there), and Chattisgarh (whose explanations were explored in the section above).

Key Finding #5: A rise in merchandise trade has surprisingly contradictory implications for formal employment of men vs women. Overall, there is a clear divide geographically between the correlation of formal work and merchandise trade.

The trend for Men



- As far as formal work for men is concerned, the map that I generated is almost a mirror image of the casual work map - which fits logically. If there is a low (negative) correlation between casual work and bulk merchandise trade for men, then there should be a high (positive) correlation between merchandise trade and formal job creation (assuming there is job creation happening overall).

- For most of India, including both rich and poor states, there is a high formal job creation for men as merchandise trade increases. The image above shows that there is no obvious north-south or east-west divide at play.

- The notable exceptions to this trend are the southern states of Karnataka and Telangana, the north-eastern states, and the belt from Himachal Pradesh-Uttarkhand-Uttar Pradesh, Jammu and Kashmir, Ladakh, and Goa.

    - In Karnataka and Telangana, the reasoning for a neutral (zero) to negative correlation could be due to the fact that these regions are very service-sector dominant and the manufacturing / transport that does take place from these states is very capital intensive (defence, pharma, aerospace, biotech). Due to this, more e-way bill generation does not lead to a strong rise in employment numbers.

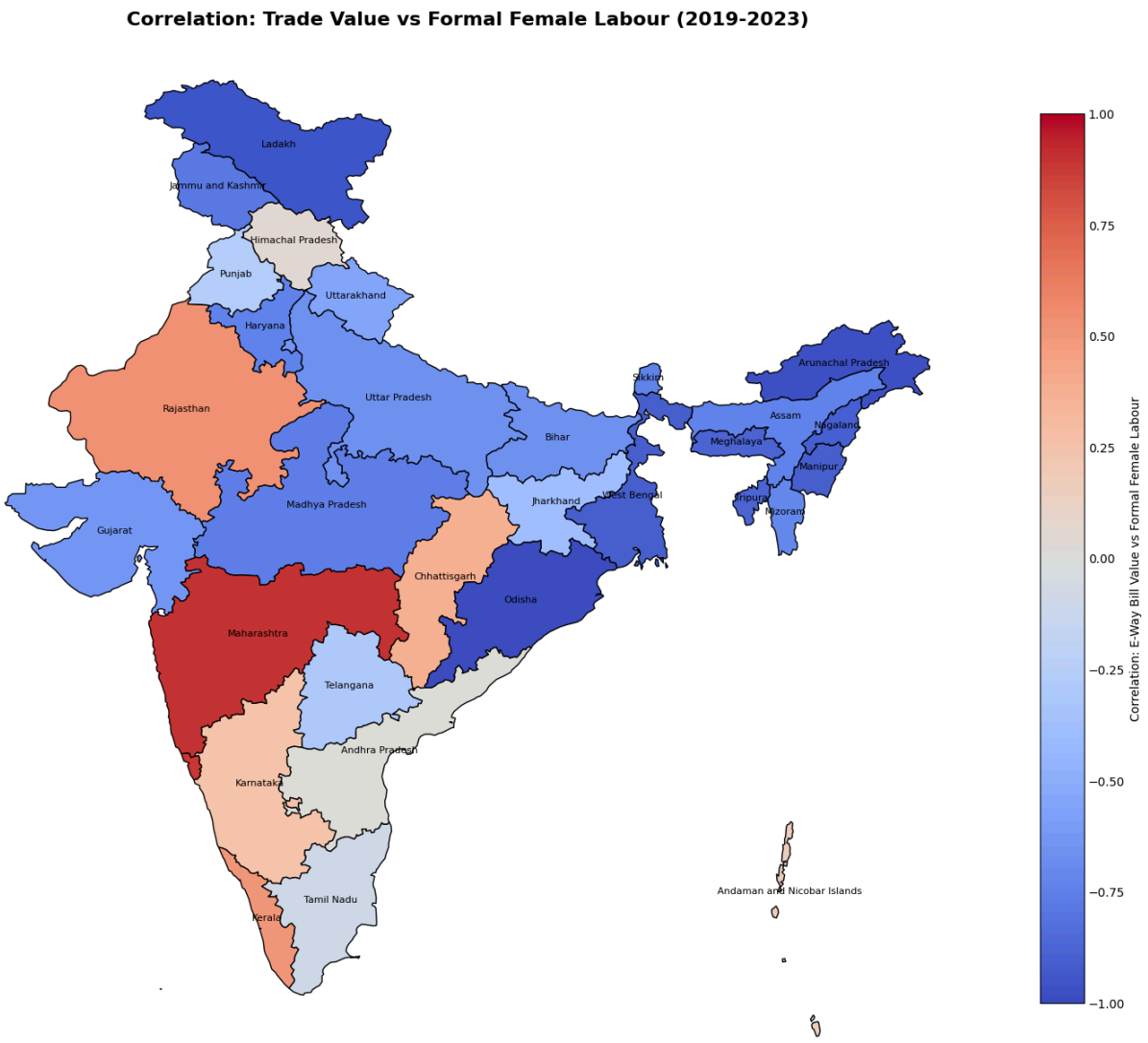
    - Uttar Pradesh is a huge manufacturing zone but the economy here has been structurally informal for a long time. This might explain some of the reason why greater e-way bill activity does not lead to a concomitant increase in formal employment.

    - Himachal Pradesh and Uttarakhand are places where high value manufacturing, particularly for the biotech and pharma sector takes place (owing to tax incentives). Moreover, it has been noted that these regions are strongly reliant on contract labour to do their manufacturing, partially explaining the neutral to negative correlation between more e-way bill activity and less formal employment.

    - The frontier states of Jammu-Kashmir and Ladakh, as well as the north eastern states, see a lot of e-way bill activity for the transport of construction equipment. This, and other merchandise trade that is occurring in these regions, does not lead to a rise of formal local employment.

    - Goa and Delhi are heavily dependent on migrant labour from the nearby regions to propel their merchandise trade increases. Thus, the inverse correlation between formal jobs and more merchandise trade.

The trend for Women



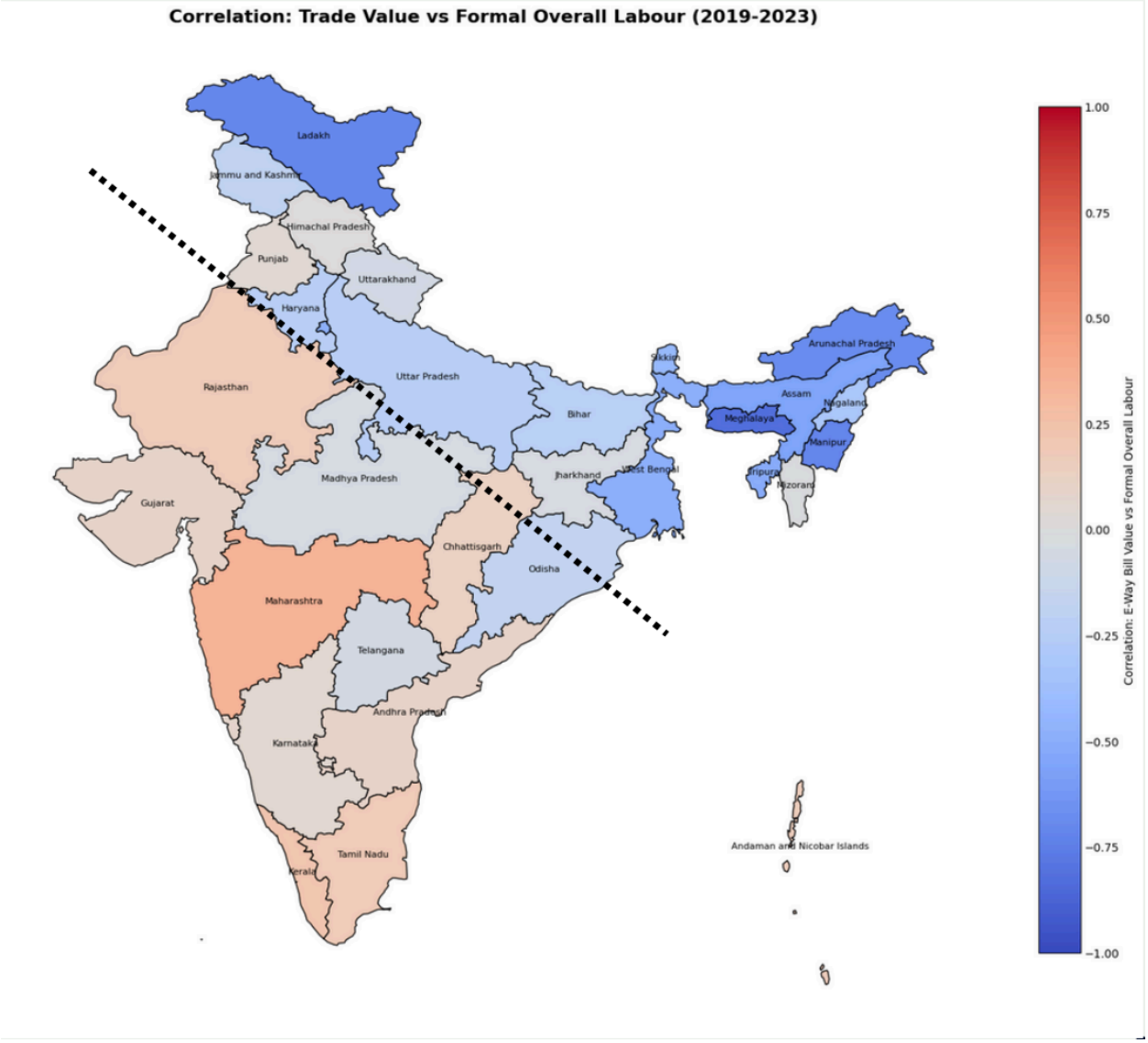
- The case of women's formal employment and its correlation with rising merchandise trade is the most counter-intuitive. While there are some states where the positive correlation holds, such as Maharashtra, Goa (opposite case for men), Karnataka, Kerala, and Rajasthan, for most of the rest of the country the opposite correlation holds, i.e., more merchandise trade leads to less formal employment for women.

- This pattern is so counterintuitive because it holds for very industrialized and non-industrialized states at the same time, and it holds for states that have higher women specific metrics (healthcare, education, etc.) as well as those with worse

indicators.

- With the distinct cultural, political, economic and security nuances for each state, I believe that this map and its implications deserves their own deep dive. No single explanation (or even 2-3 reasons) could encapsulate why we see such a widespread trend exists across the country.

The trend overall (All-Persons)



- In aggregate, for all persons in India, there seems to be a geographical divide in the country between the correlation of merchandise trade and formal employment generation. To represent this divide I have made a line on the map, where the states which lie above this line represent a negative correlation between merchandise trade and formal employment generation, while the ones below it have the opposite pattern (a positive correlation).

- This finding (about the change in correlation between one part of India versus the other) is a crucial one for the following reasons:

- There is a trend of jobless growth in India where, even if the economy is growing, high quality formal jobs are not being created at a rate that can absorb the young labour force.

- The states above the line are the ones with the highest population. It is precisely these states that should have a positive correlation if the country is to solve its employment crisis.

- The previous election was largely fought on the issue of employment and job creation in 2024. With a clear demarcation of this trend geographically, we can see where the highest percentage of people possess the concern about unemployment.

- The lack of formal job generation above the line is also a good indicator of where the migrant labour in the country is emanating from. It moves from above this line to below this line in general.

- The states that have a positive correlation are also very heterogeneous in their mix of economic fortunes and socioeconomic outcomes. Further exploration of policy direction and circumstance is needed to infer - what leads to formal employment generation? Answering this question will be key to resolving the jobless growth crisis.

Implications for the future

Based on the findings of this study, depending on the stakeholder that is analyzing the results, there can be various implications.

For policymakers, these findings offer a framework to identify states achieving desirable economic outcomes and a roadmap to begin their exploration of how to replicate those successes. Furthermore, there are also direct takeaways such as the need to address informalization of labour, the need to create more infrastructure in places where there is more e-way bill activity so that bottlenecks do not arise, etc.. Finally, policymakers can also rely on this analysis to study the outcomes of other big changes (How military infrastructure creation impacts trade and job growth, has the counter-insurgency operations in India's eastern states worked out, is there job growth following the J&K Reorganization Act of 2019, etc.)

For businesses, they can use this report and get a sense of where the economic activity is in the country, as well as the kind of labour market dynamics that exist in different geographies. This can help them guide decisions on where they would like to create new offices, where logistics hubs could be created, the level of formalization that they should expect in different states etc.

Journalists can use this analysis to report as a standalone story, or integrate it with other notable topics that they wish to explore. This includes understanding the pattern of internal labour migration, comparing states on their adoption of business friendly initiatives, exploring what kind of imbalances exist within the country, seeing which states have been a success in generating jobs - particularly formal employment for women, etc.

Finally, this analysis also provides a good starting point to initiate further deep-dives using the Government's Open Data that is available on the NDAP platform. This platform provides a lot of organized and clean data that could readily be integrated with my analysis, including external trade data, GDP data, Interest rate data, etc. Combining all of this would make a rich multi-variable analysis possible and that would be immensely helpful to all sorts of stakeholders from the government to businesses, banks, private citizens, investment vehicles, journalists, and more.

Limitations

While there have been many insights gleaned from this analysis, there are also certain limitations which restrict how much can be learnt from this data or analysis alone. Some of these limitations include:

- The E-way bill data focuses only on reporting bulk merchandise goods movement over Rs. 50,000 or around $560 in value. Thus, the smaller value goods movement (especially within a state and in the border regions) is an important vector that is not captured. This included commodities like fruits, vegetables, cereals and meat.

- To make any larger comment about a state, and the interlinkage between employment numbers & E-Way bill trends, there needs to be some integration with other aspects of economic activity as well. This includes labour migration numbers, infrastructure buildout, government expenditure (revenue vs capital), skill and educational status of the population etc.

- Most states in India are very large in size and so they should be considered as aggregates themselves. To get actionable value out of such comparative analysis (E-way bills vs Employment), a district level analysis or even smaller might be more appropriate.

- This trading data is not linked to the external economy, which is a crucial part of any trade dynamics analysis. Export-Import patterns, and the ports through which such trade passes significantly determines E-way bill activity in a region, but is not captured in this analysis.

- Higher quality employment data (not just capturing casual or salaried status) that captures the sector of economic activity and the geography of employees, along with a superimposition of that data with sector specific analysis, is a key ingredient (missing in this analysis) for making the results actionable (policywise).

- Distant regions like the Andaman and Nicobar Islands have their own dynamics, being so isolated from the mainland. Just relying on E-way bill information and employment statistics for such regions to gain insights is particularly incomplete.

Conclusion

Overall, this study has been very revealing about the interesting trade pattern dynamics that exist within the country and how merchandise trade is correlated with different forms of employment for men and women separately, as well as together.

Besides this, this project has also served as an illustration of how the new, high-frequency data of E-way bill generation could be used to gain insights about the changing patterns of merchandise trade within the country and even within states.

Correlating this with the employment status data has also provided us much new information that would be of use to a variety of stakeholders.

This analysis, I believe, serves as a good baseline for someone else to build over and add more vectors of analysis, to generate more actionable insights for businesses, policymakers, and banks. Many such directions could be taken from the limitations section above.

Finally, an aspect of the project that didn't get captured in this report but is worth mentioning, is the availability of good quality and relatively clean data at the Government of India's new data portal. It is a major step forward when it comes to transparency and accountability, and is worth praising for what it achieves (collecting and cleaning data from various agencies and surveys that have historically been scattered in various locations and never accessible under one platform for analysis). It contains a rich variety of datasets (over a thousand) that can seed their own analysis - just like this one. I look forward to seeing the 2026 Census data there once it concludes.