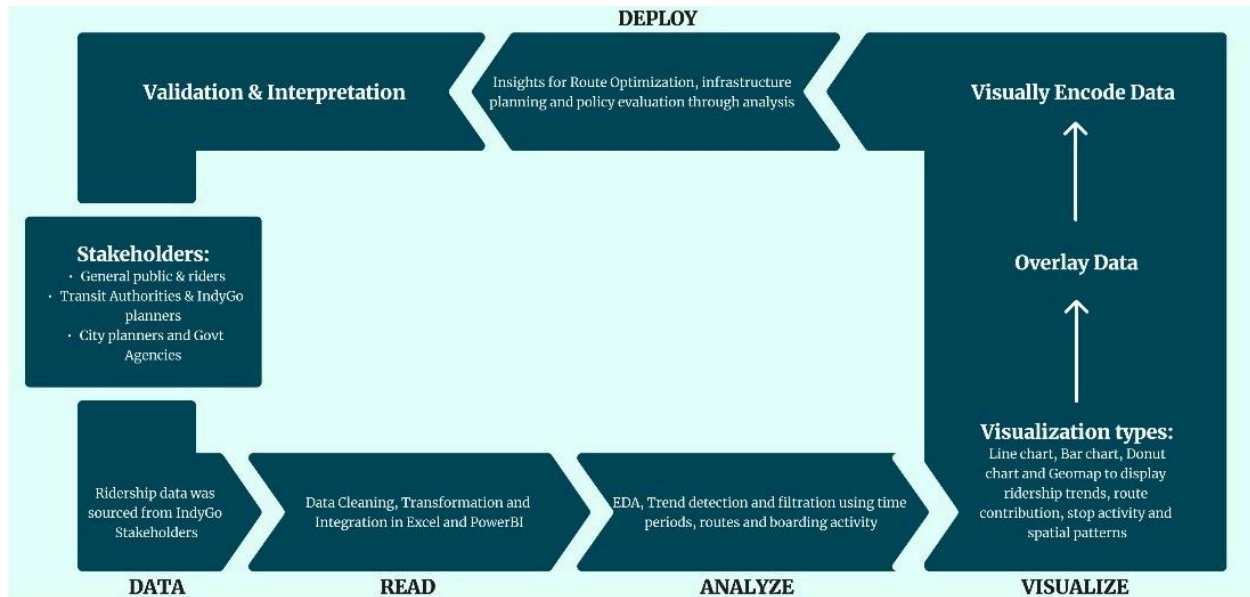# IndyGo Ridership PowerBI Dashboard

Anish R Khatavkar, Arnav Srivastava, Shruthi Nagappan, Uthra Sridhar, Rishikesh Vittal

Luddy School of Informatics, IUB

Figure 1. Visual abstract summarizing the IndyGo Ridership Dashboard project, highlighting the transportation challenges addressed, the primary data sources utilized, the data cleaning and analysis methods applied, and the key visualizations developed to support transit planning and decision-making.

1. INTRODUCTION AND PRIOR WORK

Public transportation agencies face growing challenges in providing reliable, efficient services for an increasingly diverse rider base. For IndyGo, Indianapolis's primary transit provider, understanding ridership trends is essential for evaluating system performance and guiding improvements. As transit systems evolve, having timely ridership data is critical for making informed decisions. Power BI dashboards help meet this need by tracking patterns, identifying service gaps, and driving improvements through data-driven insights.

Dashboards from other cities show both strengths and limitations. King County Metro in Seattle displays ridership and performance but lacks maps and time-specific breakdowns. RideKC in Kansas City offers ridership data but makes it difficult to compare across days or routes. MnDOT's dashboards are informative but limited in interactivity and customization, making deeper analysis challenging.

These examples show the need for dashboards that are intuitive, flexible, and user-friendly, allowing exploration by time, route, and location. This project aims to develop an IndyGo Ridership Power BI Dashboard to close those gaps, featuring route-level analysis, hourly ridership trends, and interactive maps. By leveraging Power BI's capabilities, the dashboard will promote transparency, smarter decision-making, and a more responsive transit system for Indianapolis.

1.1 Stakeholder Groups

The IndyGo Ridership Power BI Dashboard is designed to meet the needs of a wide range of stakeholders working to improve public transit. For riders and the general public, it provides easy access to ridership data, helping them plan travel more efficiently. Transit authorities and IndyGo planners use it to monitor trends, optimize routes, and make informed service decisions. City planners and government agencies rely on the data to guide infrastructure projects and promote sustainable urban mobility. Researchers and analysts use it to assess transit demand and forecast future needs, while private sector innovators and developers leverage the data for building smart mobility solutions and integrated applications. By serving these diverse groups, the dashboard helps drive smarter transit planning, supports innovation, and contributes to a more efficient and accessible transportation network for Indianapolis.

1.2 Stakeholder Needs

To effectively support its users, the IndyGo Ridership Power BI Dashboard must address the diverse needs of each stakeholder group. For the general public and riders, it offers clear ridership trends, interactive tools, and simple summaries to help plan better trips. Transit authorities and IndyGo planners rely on detailed insights to monitor ridership changes, track peak times, and optimize service and resources. City planners and government agencies use high-level reports and detailed data to guide infrastructure investments and policy development. Researchers and analysts need comprehensive datasets for trend analysis and predictive modeling, while private sector innovators and developers use transit data and APIs to create smart mobility solutions. By meeting these needs, the dashboard helps

improve public transit accessibility, supports smarter planning, and fosters innovation across Indianapolis.

2. DATA ACQUISITION

All primary data for this project was obtained directly from IndyGo stakeholders, providing a highly reliable and authentic foundation for our analysis. The dataset includes structured information collected from sources such as GPS trackers, Automatic Passenger Counters (APCs), scheduled trip records, and historical ridership data. Since the project focuses specifically on Indianapolis's transit system, we intentionally limited our dataset to these primary sources to maintain consistency, relevance, and local accuracy. Although the data came from a trusted stakeholder, we still approached it carefully, recognizing that even reliable datasets can sometimes contain inconsistencies. During the data preparation phase, we removed unnecessary columns, handled missing values, addressed duplicate records, and cross-referenced fields against the provided data dictionary to ensure that only the most essential and accurate information was retained. Basic validation steps, including consistency checks and anomaly detection, were performed to further strengthen the quality of the dataset. This thorough preparation not only improved the clarity and focus of the data but also created a transparent and scalable foundation for delivering actionable insights that support IndyGo's transit planning, service optimization, and public engagement efforts.

2.1 Data Sources

The data used for this project is the primary data source provided by the stakeholders. This includes 9 tables, with each table consisting of 3-6 relevant columns. Consistency, standardized formatting and seamless integration between all tables has been ensured as this is a primary data source.

Table Name	Num. Records	Num. Cols
DimRoute	1,236	5
DimDate	1,096	3
FactTimepointAdherence	1,923,871	6
FactHeadwayAdherence	19,787,279	5
FactSegmentAdherence	20,618,440	27
DimStop	19,349	3
DimBlock	880	3
DimUser	2,123	4
DimTrip	226,133	4

2.2 Data Description, Quality and Coverage

The primary dataset provided by the stakeholder consists of 9 tables.

- Dim Route: Each record references a route, with additional details
- DimDate: Each record references a trip date.
- FactTimepointAdherence: Each record details variance between scheduled and actual departure/arrival times.
- FactHeadwayAdherence: Each record details variance between previous departure times.
- FactSegmentAdherence: Each record details ridership metrics for each stop, departure and arrival time schedules, variances and total trip duration and variances.
- DimStop: Each record references a particular stop.
- DimBlock: Each record references a particular block.
- DimUser: Each record details operator information.
- DimTrip: Each record details tripID and corresponding details.

The dataset operates at the "stop" level of granularity, which is the most detailed view available. This level of detail enables highly specific insights into performance and ridership trends at individual bus stops. Covering a span from 2022 to 2024, the dataset supports robust year-over-year comparisons, seasonal pattern analysis, and monthly trend identification. While the data was sourced directly from stakeholders and is highly comprehensive, standard validation and cleaning processes were still necessary to address occasional inconsistencies and ensure the quality of insights derived from it.

3. DATA ANALYSIS

To begin working with the primary dataset, we used Excel and Google Sheets to explore the data structure, assess the overall scale, and perform initial cleanup. Pivot tables were especially helpful during this stage, allowing us to quickly summarize large volumes of records and identify unnecessary columns that could be removed. This early exploration helped simplify the dataset and ensured that only the most relevant fields were carried forward for deeper analysis and dashboard development.

3.1 DATA CLEANING

We focused specifically on refining the FactSegmentAdherence table to address gaps and inconsistencies at the trip level. For each trip record, we carefully verified that the associated route key matched the correct coverage group. When missing or unclear coverage details were found, we created a new column within the FactSegmentAdherence table to explicitly capture the assigned service coverage for that trip. As part of the data cleaning process, we also handled NULL values by converting them to 0 and then changing the column type to Decimal Number, following a standardized procedure applied across multiple columns. These careful adjustments helped ensure that the dataset could be reliably filtered, grouped, and analyzed during dashboard development. By the end of the cleaning phase, we had established a strong and accurate foundation that was essential for producing trustworthy and actionable insights.

3.2 DATA PROCESSING

After completing the cleaning phase, we moved into processing the dataset to make it even more useful for time-based and operational analysis. One important transformation involved creating a new "Depart Time Hour" column by extracting just the hour component from each trip's departure timestamp. This allowed us to break down ridership patterns by different parts of the day, such as morning rush hours, midday lulls, and evening commutes, making it easier to visualize and interpret daily transit behaviors.

Additionally, to better capture stop-level activity, we developed a more accurate boarding metric by analyzing Departure Boards data[AK1]. By creating a new field to measure the number of boardings at each stop, we were able to gain a much clearer and more detailed understanding of where passenger demand was highest throughout the day. This enhancement allowed for more targeted insights into stop performance across the transit network.

We also improved the usability of the DimRoute table by combining two columns RouteFareboxID and RouteInternetName [AK1] into a single, simplified route identifier that was more intuitive for dashboard users. To maintain focus on public-facing transit services, we filtered out internal or non-public route[AK2] s (for example. Such as Deadhead and Special Routes.) that were not relevant to the broader analysis. As we transitioned into building visualizations, we created several custom measures in Power BI tailored to different analytical needs. Some of these measures were experimental and helped us explore ideas during development, while others were finalized and directly incorporated into the dashboard depending on their relevance and contribution to the overall story we wanted to tell. Together, these processing steps allowed us to turn a large, complex dataset into an accessible and insightful foundation for the dashboard.
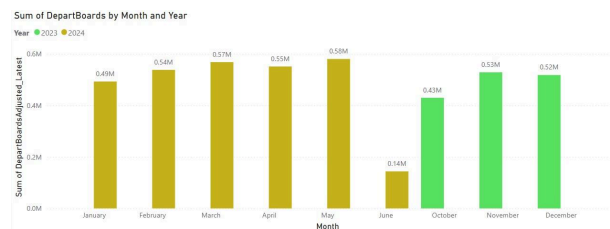
4. VISUALIZATIONS

The dashboard uses different types of charts and varieties of slicers for the users to interact on each page to show ridership patterns accurately. On the Overview page, a bar chart that shows total ridership and daily boardings by month, and also highlights which route has the most riders. In the Route Performance page, a donut chart that shows how boardings are divided among different types of routes, like Frequent, Basic, Rapid 90, and Coverage. The Hourly Performance page uses a stacked area chart to show average boardings by day and a slicer to explore by different hours of the day, helping find peak hours. The Geospatial Analysis page has a map that shows where bus stops are located, with bigger circles showing stops with more people getting on. Finally, the Coverage Analysis page includes1 tables rank top routes by ridership within different coverage groups, and another donut chart summarizes the overall distribution of ridership across service types. The entire dashboard helps IndyGo planners and the public better understand ridership patterns and make better decisions.
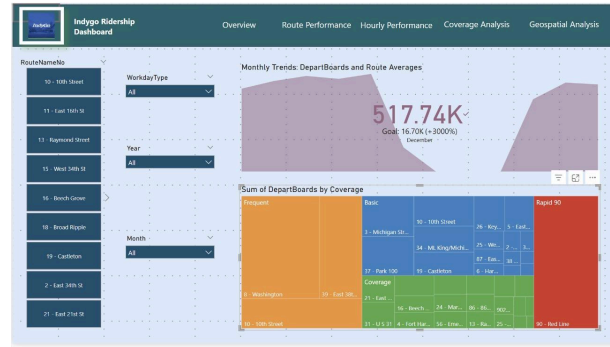
Figure 2. IndyGo Ridership Power BI Dashboard: Route Performance and Coverage Analysis

4.1 BAR CHART VISUALIZATION

This bar chart shows the total number of boardings on each month for the years 2023 and 2024. Each bar represents one month, and the height of the bar shows how many people boarded buses during that month. Yellow bars are for the year 2024, and green bars are for 2023. You can easily see which months had higher or lower ridership. For example, May 2024 had the highest boardings, while June 2024 had much fewer. This information can help transit authorities make informed decisions to optimize service, allocate resources, and improve the overall efficiency of the Indianapolis public transportation network.

Figure 3. Monthly Ridership Trends (Sum of DepartBoards) by Month and Year.

### 4.2 AREA CHART VISUALIZATION

This stacked area chart shows the average number of boardings for each day of the month. The x-axis represents the days from 1 to 31, and the y-axis shows the average number of people boarding buses. The height and shape of the green area highlight how boardings vary across the month. You can see that some days, like the 11th (0.231), 25th (0.229), and 31st (0.238), have higher average boardings, while days like the 7th (0.136), 14th (0.130), and 21st (0.130) show lower averages. Additionally, the KPI displayed at the top right corner shows the hourly average of DepartBoards for all routes combined, allowing users to easily compare the performance of a selected route against the overall network average. Together, the area chart and KPI help identify patterns in daily ridership and highlight how individual routes perform relative to the system as a whole.
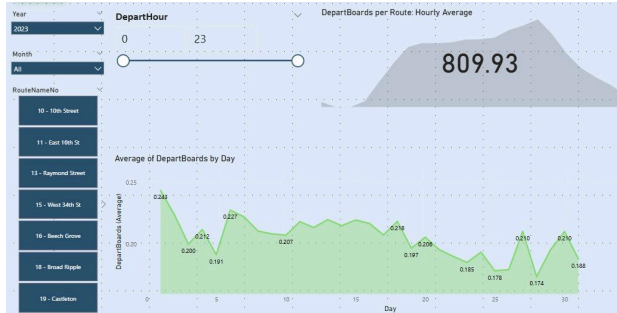
Figure 4. Daily Average Ridership Trends (Average of DepartBoards by Day).

4.3. GEOSPATIAL VISUALIZATION

This map shows the total number of boardings at different bus stops across Indianapolis, highlighting the busiest stops on the network. Each dot represents a bus stop, with the size of the dot indicating the number of boardings larger dots represent stops with heavier rider activity. The map clearly shows that most of the largest circles are concentrated in downtown Indianapolis, with Pennsylvania St & South St emerging as one of the busiest locations with 175.32K total boardings. To better emphasize this, we synced the filters across the dashboard, allowing users to observe all route stops together and see how boardings are distributed citywide. This visualization makes it easy to identify areas with the highest transit usage and understand where public transportation is most heavily relied upon.
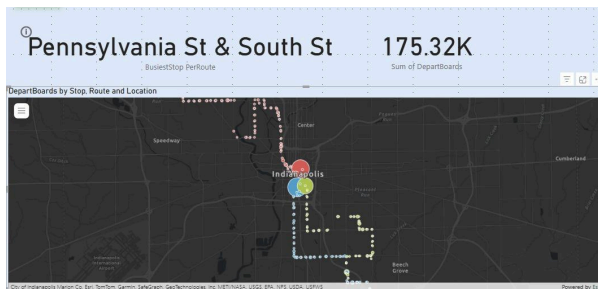
Figure 5. Bus Stops by Route and Location in Indianapolis

5. USAGE AND CRITIQUE OF AI TOOLS

AI was used to help structure stakeholder needs and assisted in understanding where improvements can be made to existing visual ideas. It was useful in helping us choose between color-blind friendly themes and visual styles. It also helped by giving us certain tool suggestions to complete certain visuals on our dashboard, although they were not used as the size of the dataset was not compatible with said tools. We believe that this is where AI lacks the judgemental skills possessed by humans, where certain obvious results and ideas must be ruled out because of certain limitations. Overall, AI has been of help to us in this assignment and we believe that it's role will only grow in the future.

6. INTERPRETATION OF RESULTS

The updated dashboard offers a clear and interactive view of ridership patterns across Indianapolis, helping planners and stakeholders make more informed decisions. The Overview page highlights total ridership trends and key routes by month, while the Route Performance page shows how boardings are divided among service types like Frequent, Basic, Rapid 90, and Coverage routes. The Hourly Performance page tracks daily and hourly boarding trends, helping identify peak times. The Geospatial Analysis map highlights the busiest stops, with larger circles showing heavy rider activity around downtown, especially at Pennsylvania St & South St. The monthly ridership bar chart compares 2023 and 2024 patterns, revealing seasonal shifts like the rise in May and drop in June. Together, these insights support smarter service planning, better resource allocation, and efforts to make Indianapolis's public transit system more efficient and accessible.

REFERENCES

1. K. Börner. *Atlas of Knowledge: Anyone Can Map*. MIT Press, 2015.
2. K. Börner, A. Bueckle, and M. Ginda. "Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments." *PNAS*, 116 (6): 1857-1864, 2018.
3. King County Metro Transit. (n.d.). Metro Transit official website. Retrieved April 2025, from https://kingcounty.gov/depts/transportation/metro.aspx
4. RideKC. (n.d.). Public Transit Dashboard. Retrieved April 2025, from https://ridekc.org/planning/dashboard
5. Minnesota Department of Transportation (MnDOT). (n.d.). Transit Ridership Measures. Retrieved April 2025, from https://www.dot.state.mn.us/measures/transit-ridership.html
6. Chicago Transit Authority (CTA). (n.d.). Ridership Reports and Data. Retrieved April 2025, from https://www.transitchicago.com/ridership/#prepared.