

**DATA VISUALIZATION FOR ANALYTICS**

*DAAN- Data Visualization Course Project*

*Tushar*

[*Tpt5409@psu.edu*](mailto:Tpt5409@psu.edu)

**School of Graduate Professional Studies**

MPS/MS in Data Analytics

November, 2025

# Document Control

## Work carried out by:

|  |  |  |
| --- | --- | --- |
| **Name** | **Email Address** | **Other** |
| **Tushar** | **tpt5409@psu.edu** |  |
|  |  |  |
|  |  |  |

## Revision Sheet

|  |  |  |
| --- | --- | --- |
| **Release No.** | **Date** | **Revision Description** |
| 1 | 11/8/2025 | Version 2 |
| 2 | 11/9/2025 | Version 3 |
| 3 | 11/16/2025 | Version 4 |
| 4 | 11/23/2025 | Version 5 |
| 5 | 12/07/2025 | Version 6 |
|  |  |  |
|  |  |  |

**Note**: For the project assignment deadline submission refer to the syllabus. Each week’s assignment is for 10 points and the project assignment is worth 25% of your final grade. The project presentation is worth 5% of your final grade.

**TABLE OF CONTENTS**

1. **Business Goal -------------------------------------------------------- I**
2. **Exploratory Data Analysis ----------------------------------------- II**
3. **Purpose Of Dashboard & Target Users ------------------------ III**
4. **Design Principles and Techniques ------------------------------ IV**
5. **Final Dashboard ------------------------------------------------------ V**
6. **Conclusion-------------------------------------------------------------- VI**

**General Guidelines**

1. Please use this template document to complete each deliverable assignment.
2. Each assignment must be submitted by the due date in the Course Schedule.
3. All figures should be followed by a brief description about the figure.
4. Figures can be hand-drawn and scanned in some circumstances, but the hand-drawn figure should be clear and legible to obtain full credit. Unclear hand-drawn figures will receive partial credit. For constructing figures and diagrams it is advised to use tools.
5. Figures and tables should have appropriate captions. For documenting and referencing styles please follow the APA or MLA writing style.
6. Please make sure that you provide a reference section.
7. Any material text or figure taken from books, journals or Internet should be referenced. If you have a sentence or a figure that does not belong (authorship) to you, they must be clearly referenced. If you fail to do so your report will be considered as a case for plagiarism. It is your duty to make sure that your report is free from any activity related to plagiarism. Please see the section on Academic Integrity found below. The penalty for plagiarism will be a “0” awarded to your report. Thus, it is good to keep it simple, always have the principle to acknowledge people for their contributions.

#### **Academic Integrity**

Academic integrity — scholarship free of fraud and deception — is an important educational objective of Penn State. Academic dishonesty can lead to a failing grade or referral to the [Office of Student Conduct](http://www.sa.psu.edu/ja/).

Academic dishonesty includes, but is not limited to:

* cheating
* plagiarism
* fabrication of information or citations
* facilitating acts of academic dishonesty by others
* unauthorized prior possession of examinations
* submitting the work of another person or work previously used without informing the instructor and securing written approval

In cases where academic integrity is questioned, [procedure requires an instructor to notify a student](http://www.psu.edu/oue/aappm/G-9-academic-integrity.html) of suspected dishonesty before filing a charge and recommended sanction with the college. Procedures allow a student to accept or contest a charge. If a student chooses to contest a charge, the case will then be managed by the respective college or campus Academic Integrity Committee. If a disciplinary sanction also is recommended, the case will be referred to the [Office of Student Conduct](http://www.sa.psu.edu/ja/title=).

All Penn State colleges abide by this policy, but review procedures may vary by college when academic dishonesty is suspected. Information about Penn State's academic integrity policy and college review procedures is included in the information that students receive upon enrolling in a course.

Penn State students are expected to act with civility and personal integrity; respect other students' dignity, rights, property, and help create and maintain an environment in which all can succeed through the fruits of their own efforts. An environment of academic integrity is requisite to respect for oneself and others, and a civil community.

#### For More Information on Academic Integrity at Penn State

Please see the [Academic Integrity Chart](http://www.campuses.psu.edu/CAO.pdf)  for specific college contact information or visit one of the following URLs:

* Penn State Senate [Policy on Academic Integrity](http://www.psu.edu/dept/oue/aappm/G-9.html)
* [iStudy for Success!](http://istudy.psu.edu/tutorials/) — learn about plagiarism, copyright, and academic integrity through an educational module
* [Turnitin](http://tlt.its.psu.edu/turnitin) a web-based plagiarism detection and prevention system.
  1. **Defining Business Goal for the Dashboard**

This dashboard's main objective is to give a clear, data-driven picture of how the market for Alzheimer's treatments is changing over time. The dashboard helps identify growth trends, changing prescriber behavior, and new competitive opportunities by monitoring Total TRx volumes across brands, specializations, and channels. Giving commercial, marketing, and strategic teams a single, interactive tool to track brand performance, pinpoint high-value physician groups, and comprehend how TRx distribution varies across channels is the goal.

In the end, the dashboard makes it possible to make quicker, better judgments on where to concentrate marketing efforts, how to distribute resources, and which market segments have unrealized potential.

**Business Question Which My Dashboard Should Answer  
Question 1**

**How have Total TRx volumes for key Alzheimer’s therapies changed over time, and which brands are gaining or losing market momentum relative to competitors?**

This inquiry looks at the monthly variations in Total TRx volumes for the main Alzheimer's treatments. By comparing brand trajectories side by side, the dashboard enables users to rapidly determine which goods are gaining traction and which are losing market share. This aids stakeholders in keeping an eye on changes in the competition and assessing the effects of market events, new product launches, or advertising campaigns.

**Question 2:**

**Which physician specialties are driving growth in Alzheimer prescriptions over time, and how is each specialty’s share of total TRx shifting month by month?**

The goal of this inquiry is to determine which medical specializations are most responsible for the rise in prescriptions. The dashboard makes it simple to spot growing or shrinking prescriber groups by displaying how each specialty's TRx share varies over time. This data helps establish whether promotional initiatives are connecting with the appropriate consumers and supports specialty-level targeting tactics.

**Question 3:**

**Which specialty channel combinations drive the highest TRx, and where are there untapped channel opportunities for key specialties?**This inquiry delves deeper into the ways that different specialties engage with distinct TRx channels. The dashboard illustrates current strengths and unexplored prospects by displaying specialty-channel combinations that produce the largest volumes. By using this information, teams can make sure that outreach initiatives are in line with the highest-potential combinations and optimize channel strategies.

**Question 4:**

**How is the TRx channel mix evolving over time, and which channels are gaining or losing share in the Alzheimer market?**

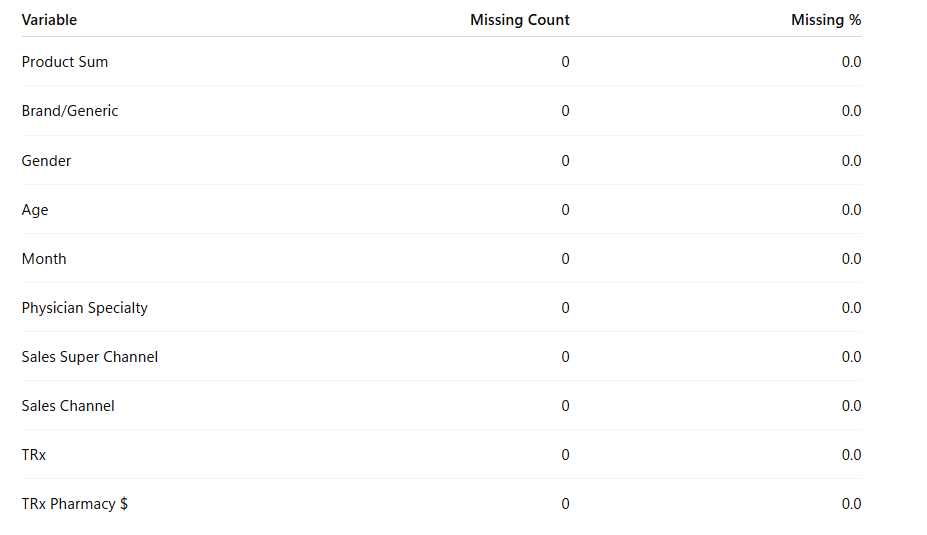
This query analyzes how the TRx channel distribution changes over time for the Alzheimer’s market. The dashboard reveals which channels are gaining traction and which are declining. Understanding channel mix evolution helps predict shifts in prescribing behavior, determine where future investments should be directed, and evaluate whether the market is moving toward retail, mail order, long-term care, or specialty channels.

**Understanding of the Provided Data Source to Address Business Goal**

The dataset offers comprehensive monthly prescription activity for Alzheimer's treatments, including Total TRx volumes split out by channel, product, and prescriber specialization. Strong trend analysis is made possible by the fact that it covers a number of years and features both well-known and up-and-coming businesses. Every entry in the dataset offers structured data that directly supports the business objective of tracking market dynamics, such as product name, time period, Total TRx count, specialized category, and dispensing channel. Building interactive visualizations that record trends, segment performance, and channel changes over time is made possible by this data foundation. Each of the four main analytical issues is supported by the dataset's structure and granularity, which allow for a thorough understanding of the Alzheimer's TRx ecosystem.

* 1. **Exploratory Data Analysis**

**Missingness**



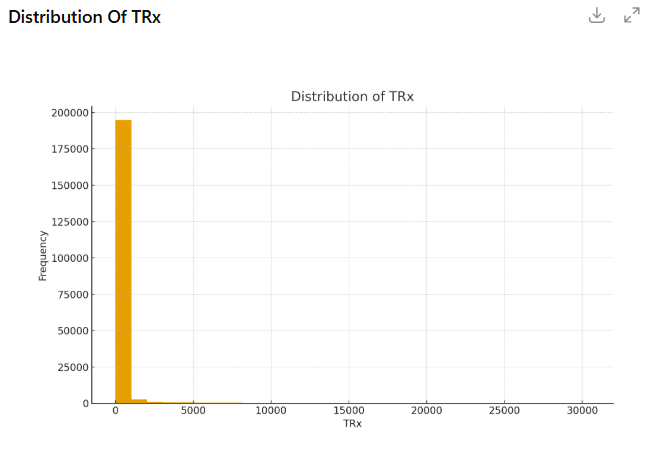
No column has any missing values (NA/blank) at the system level.  
However, several records in Gender and Age have the label "UNSPECIFIED" in the data. Instead of treating these as actual levels of the variable, I treated them as coded missing values (unknown/unspecified category).

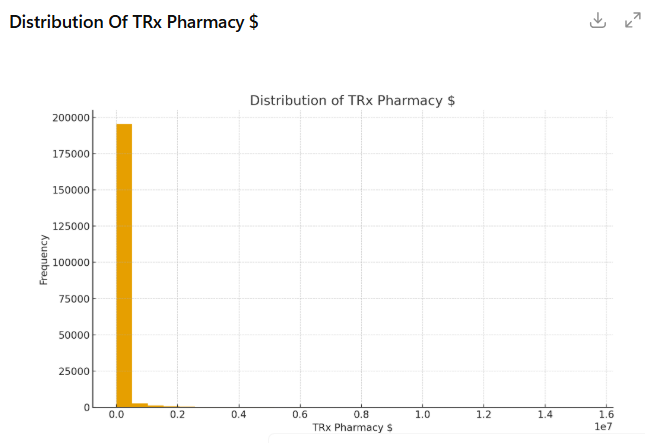
Because their absence is probably connected to patient characteristics or data collection procedures, these "UNSPECIFIED" items behave similarly to Missing Not at Random (MNAR).  
Instead of intentionally imputing or removing those entries, I maintain "UNSPECIFIED" as a separate category for the dashboard so that users can see how much of the data has unknown demographics.  
  
I don't need to perform formal imputation because the dataset is generally quite full. The sole "missingness" is already encoded as "UNSPECIFIED" and is categorical.

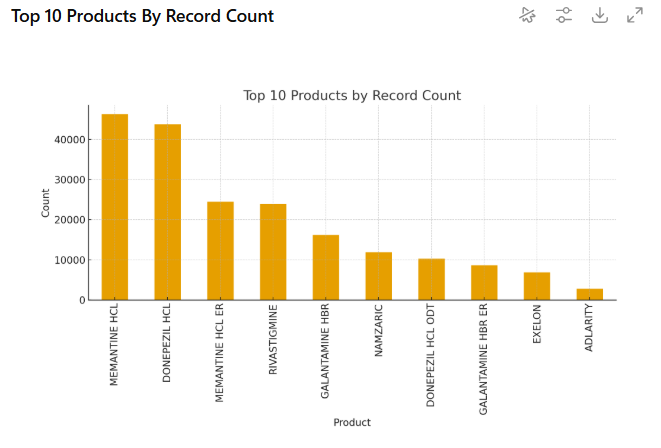
**Shape of the dataset**

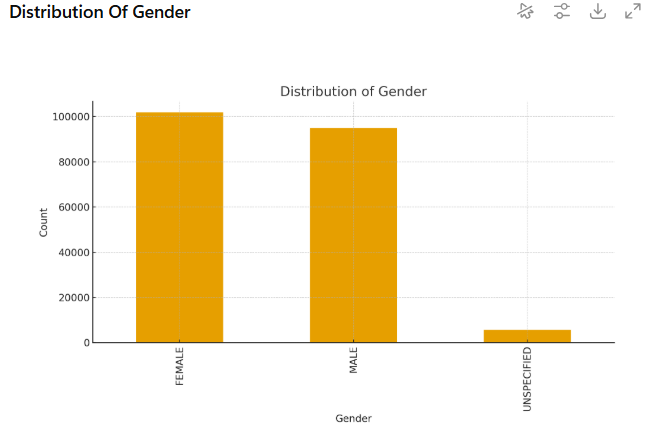
**10 columns and 202,499 rows**  
  
202,499 distinct prescription-level data, each with a distinct product, specialty, channel, demographics, and month combination.  
Ten factors, including revenue, sales channel, prescriber specialty, product details, demography, and TRx.

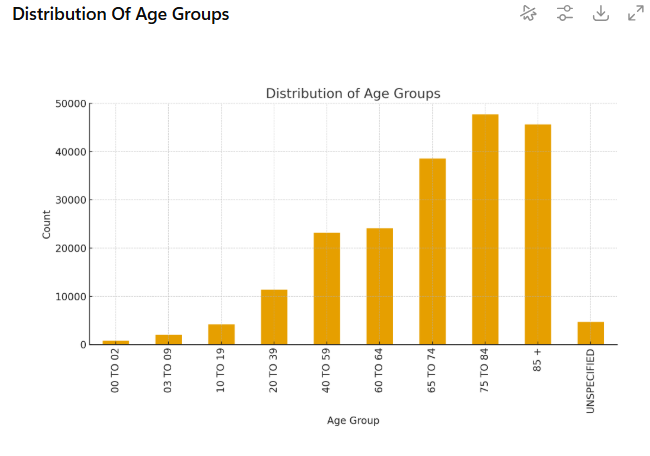
**Distributions**

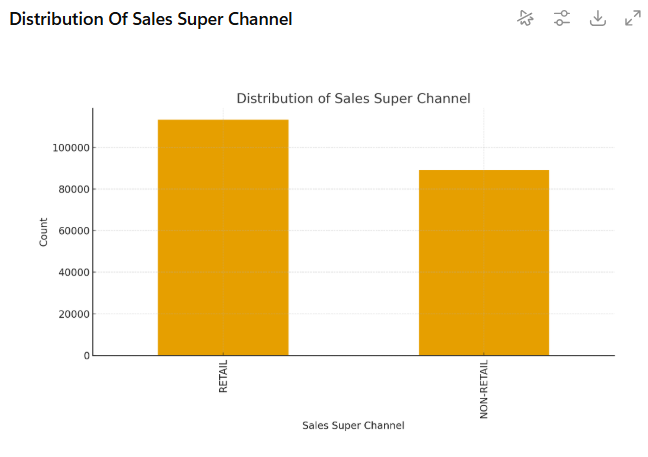












Two numerical variables

TRx – total prescriptions

TRx Pharmacy $ – total pharmacy dollars tied to those TRx

**TRx**  
The distribution has a strong right skew.With a very long tail where certain product–segment–month combinations have very high TRx counts, the majority of rows have relatively tiny TRx values (e.g., 1–100).This implies that a significant portion of the total volume is driven by a limited number of "power" combinations.

**TRx Pharmacy $**

The right skew of TRx Pharmacy $ is significantly more pronounced.  
There are a few really large revenue observations and a lot of low-revenue data.  
This pattern is anticipated as revenue is influenced by price and product mix in addition to being proportionate to TRx.

**Categorical Variables**

The dataset's categorical variables show distinct trends across channels, specializations, demographics, and products. Established Alzheimer's treatments including MEMANTINE HCL, DONEPEZIL HCL, MEMANTINE HCL ER, and RIVASTIGMINE dominate the product distribution, with newer or less commonly prescribed medications appearing in considerably lesser quantities. With a few records marked as UNSPECIFIED, the gender variable reveals a virtually equal distribution of FEMALE and MALE patients. With the bulk of prescriptions concentrated in the 65–74, 75–84, and 85+ categories, the distribution of age, which is recorded in ordered bands, reflects the anticipated demographic profile for Alzheimer's disease and very little representation in younger age groups. A wide range of prescribers can be found in Physician Specialty, with the most common occupations being Neurology, Internal Medicine, Family Medicine, and Nurse Practitioners. Important insights are also provided by channel-related variables: While more specific Sales Channel categories, like mixed retail, mail order, and long-term care, capture the many dispensing paths utilized in the Alzheimer's market, RETAIL and NON-RETAIL represent the two main high-level segments. When taken as a whole, these category distributions provide a clear picture of how prescription activity is divided among different products, patient demographics, provider types, and dispensing channels.

**Data Types and Measurement Scales**



**3. Purpose Of Dashboard & Target Users**

By combining product-level TRx trends, specialty prescribing patterns, and channel behaviors into a single, cohesive analytical tool, this dashboard aims to present a transparent, interactive, and data-driven picture of the Alzheimer's prescription market. The dashboard is intended to assist users in rapidly identifying changes in the industry, including which brands are gaining or losing traction, how specialized prescription trends change over time, and how the mix of sales channels varies throughout the market. The dashboard facilitates better strategic and operational decision-making by converting complicated prescription data into clear visual insights.

**The Dashboard's intended users**  
  
This dashboard is designed for a range of stakeholders from strategy, analytics, and commercial teams:  
  
**- Marketing teams and brand managers** should keep an eye on brand performance, assess the effectiveness of campaigns, and spot possibilities or threats from the competition.  
  
**- Sales Leadership & Field Strategy Teams:** To determine which channels and specializations generate the highest TRx and to improve targeting tactics.  
  
**- Market Access and Channel Strategy Teams:** To examine the differences in prescribing practices between mail order, retail, non-retail, and long-term care channels.  
  
**- Data analysts and business intelligence teams:** To enable more in-depth investigation of TRx behaviors, verify theories, and produce insights for partners across functional boundaries.  
  
**- Executive Stakeholders:** To assess market changes and high-level trends for strategic planning.

**Data Driving Each Visualization**

Each component of the dashboard is powered by specific fields in the Alzheimer’s dataset:

**- Product TRx Trends Over Time**

* **Data used:**  
  Product Sum, Month, TRx
* **Purpose:**  
  To compare market performance of key Alzheimer’s therapies and identify upward or downward TRx momentum.

**- Specialty Contribution & Growth**

* **Data used:**  
  Physician Specialty, Month, TRx
* **Purpose:**  
  To show which specialties contribute most to overall TRx and how their shares shift month-to-month.

**- Specialty × Channel TRx Combinations**

* **Data used:**  
  Physician Specialty, Sales Channel / Sales Super Channel, TRx
* **Purpose:**  
  To highlight high-performing specialty–channel combinations and uncover underutilized channel opportunities.

**- Channel Mix Evolution**

* **Data used:**  
  Sales Channel, Sales Super Channel, Month, TRx
* **Purpose:**  
  To track how the prescription mix across retail, mail order, and non-retail settings changes over time.

**Analytical Questions Users Can Respond to on the Dashboard**  
  
Four main analytical issues that represent the flow of prescriptions through products, specializations, and channels are intended to be addressed by the dashboard:   
  
**Which products are gaining or losing market momentum in comparison to rivals, and how have Total TRx volumes for important Alzheimer's medicines altered over time?**  
- Monthly comparisons of competing performance and brand trajectories are available to users.   
  
**Which physician specialties are driving growth in Alzheimer’s prescriptions over time, and how is each specialty’s share of total TRx shifting?**   
  
- This makes it easier to determine which specialty respond best to treatment trends and marketing initiatives.

**Where are there unexplored channel prospects for important specializations, and which specialty-channel pairings produce the highest TRx?**   
  
- Users are able to identify gaps and high-value niches where prescription potential has not yet reached its full potential.   
  
**Which channels are gaining or losing market share in Alzheimer's, and how is the TRx channel mix changing over time?**   
  
- This aids channel and market access teams in determining where prescriptions are going and how to modify their approach.

* 1. **Design Principles and Techniques**

**Tool Used and Version**

I utilized ChatGPT (OpenAI GPT-5.1) as the one of the tools for this assignment to help with the construction of analytical framework, visualization methods, and dashboard design concepts. The business questions were refined, suitable visual encodings were found, design techniques were compared, and narrative explanations for the report were produced with the help of ChatGPT. The ChatGPT Plus platform, GPT-5.1 model version, which provides sophisticated reasoning, data-analysis guidance, and dashboard design recommendations, was used for all interactions.

**Suggested List of Design Principles & Techniques + Prompts Used**

I utilized ChatGPT to create an external, expert-informed list of suggested visualization principles and methodologies in order to improve the dashboard's visual appearance and make sure my selected principles suitably fit with the four analytical questions. The reasoning behind the concepts I chose for each analytical inquiry was strengthened and validated by these suggestions. The prompts I used and the matching "Suggested List" that ChatGPT offered are listed below.

**Prompts Used to Query ChatGPT**

1. “What visualization design principles should I use for time-series TRx analysis across multiple products?”
2. “Suggest appropriate visualization techniques for analyzing specialty-level TRx contributions over time.”
3. “What principles help effectively display specialty × channel combinations and highlight patterns clearly?”
4. “What visualization methods best show how prescription channel mix evolves over time?”
5. “Provide design principles from Stephen Few, Gestalt, and Tuftte that are most suitable for dashboard analysis with TRx, channel mix, and specialty contributions.”
6. “Recommend techniques that enhance clarity, reduce cognitive load, and support fast pattern recognition in dashboards.”

**Suggested List of Design Principles & Techniques (from ChatGPT)**

The following broad visualization design ideas and strategies were suggested by ChatGPT based on the aforementioned prompts. These naturally correspond with the analytical requirements of Questions 1-4 and are divided into (a) Design Principles and (b) Visualization Techniques.

**A.** **Recommended Design Guidelines**

**1. The Principles of Stephen Few**  
  
- Avoid ornamentation and distracting features and use clear visual encodings that convey no more or less than what is important in the message.  
  
- Steer clear of visual distinctions (shapes, colors, patterns) that don't accurately reflect variations in the data.  
  
- Since length and 2-D location provide the maximum perceptual accuracy, utilize them as the main encoding for quantitative variables.  
  
- To lessen cognitive burden and encourage chunking in short-term memory, provide relevant comparisons in a single eye span.  
  
- Make sure that any difference in encoded values is proportionate to the differences in underlying data.

**2. Gestalt Concepts**- To help users quickly identify conceptual clusters, apply the Principle of Enclosure to visually group related categories (such as specialties, age bands, and channels).  
  
- To organize complex multivariate displays (such specialty × channel combinations), use closeness, similarity, and enclosure.

* 1. **Tufte's Graphical Integrity Principles**- To prevent confusion, make sure all visualizations have precise labeling, axis explanations, and annotations.

- To preserve integrity, simplicity, and clarity, remove any superfluous chart-ink.  
  
- To prevent misunderstandings, use standardized or inflation-adjusted units for time-series incorporating monetary variables.

- To prevent the graph from overstating or understating trends, maintain proportionate visual depiction.

**4. Principles of Cognitive and Interaction Design**  
- To preserve semantic meaning, use color consistently (e.g., fixed colors for items, specialties, or channels).  
  
- Incorporate interactive filtering to enable users to examine intricate patterns and lessen clutter.  
  
- For more in-depth understanding without overpowering the image, include tooltips, highlights, and annotations.

**B. Recommended Visualization Methods**The following methods were suggested by ChatGPT as the most suitable for the Alzheimer's TRx analytics:  
  
- TRx time-series trend line charts for several Alzheimer's treatments (Analytics Question 1).  
  
- To see how the channel mix has changed over time, use stacked area charts (Analytics Question 4).  
  
- Monthly comparisons of specialty-level TRx contributions using grouped or stacked bar charts (Analytics Question 2).  
  
- Heatmaps for identifying high-value clusters by examining specialty × channel TRx volumes (Analytics Question 3).  
  
- To compare trends across goods or specialties without overcrowding a single chart, use small multiples. To facilitate user research, there are interactive filters for Month, Product, Specialty, and Channel. Contextual detail can be enhanced with hover-based tooltips without creating clutter. Annotations to draw attention to data spikes, turning points, or new patterns.

**Comparison Between Proposed vs. Suggested Design Principles & Techniques**

The suggested list produced by ChatGPT and my suggested design principles, which were chosen especially for Analytics Questions 1 through 4, show both strong alignment and significant differences that affected the final layout of my dashboard design. While clarity, perceptual correctness, and cognitive efficiency are generally emphasized in both lists, the proposed principles broaden the design approach by include more sophisticated interaction features, broader visualization tactics, and other cognitive-science considerations.

My suggested design mostly based on Gestalt grouping, Tufte's graphical integrity rules, and Stephen Few's ideas across all four analytical issues. These ideas were selected because they offer a solid theoretical foundation for genuine, accurate, and cognitively effective visual representations. For instance, the time-series, ranking, and comparison analyses used in TRx dashboards are well-aligned with principles like displaying only what is pertinent, making sure visual differences match actual data differences, and encoding quantitative values using length and position. This theoretical underpinning aligns with the principles proposed by ChatGPT, which also stressed decreasing non-data ink, avoiding clutter, and maintaining proportionality in all visual encodings.

Nevertheless, ChatGPT's recommendations went beyond the theoretical ideas to incorporate more practical visualization methods and more comprehensive dashboard design concerns. For example, the recommended list highlighted interaction features like filters, tooltips, and annotations all of which improve user engagement and enable users to extract deeper insights without overwhelming the interface while my principles concentrated on the integrity and perceptual accuracy of the visuals. Although they weren't specifically on my suggested list, these methods have a lot of useful applications in an interactive dashboard setting.

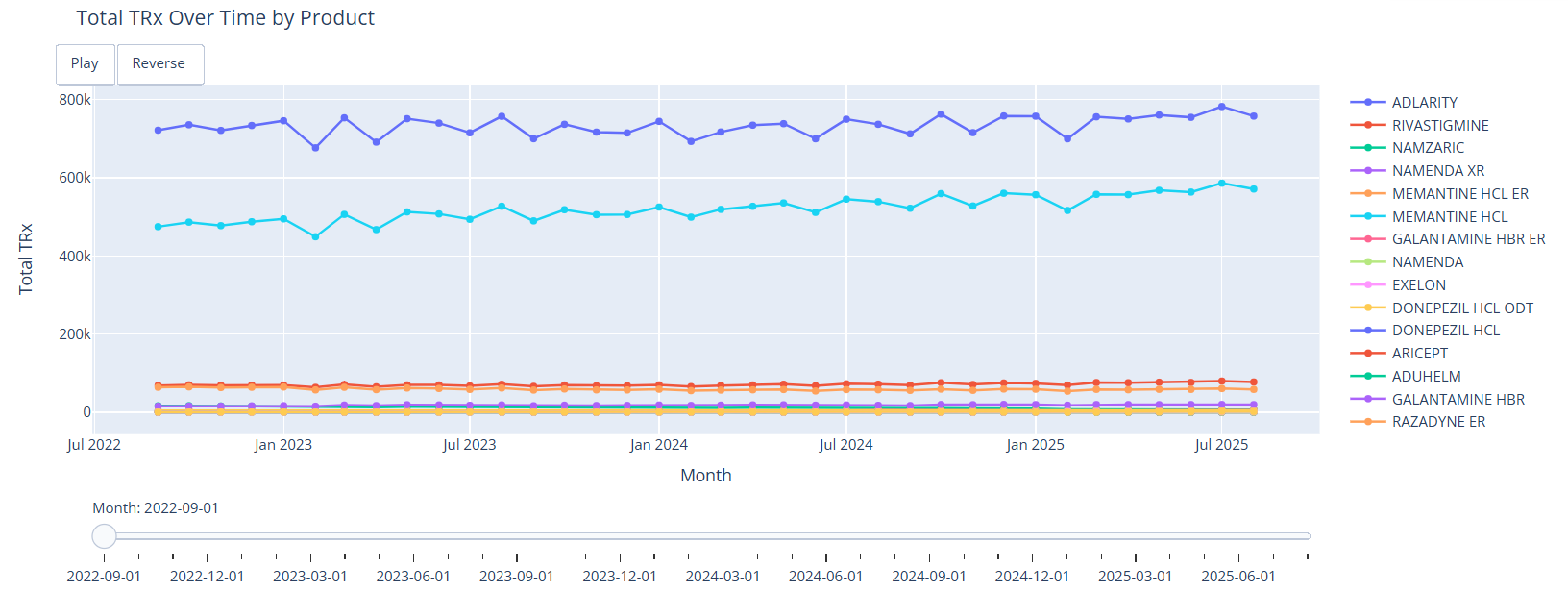
Another significant distinction is that ChatGPT suggested heatmaps, small multiples, and stacked area charts as additional visualization formats appropriate for time-series decomposition and specialist channel analysis. Although the principles I chose (such as Stephen's Principle 7 on enhancing short-term memory by providing all comparisons within eye span) completely support their use, these chart kinds were not specifically included in my initial recommended list. As a result, the recommended methods expand the visual toolset that may be used to answer the business problems without going against the fundamental ideas.

Additionally, there were significant parallels: in line with Tufte's recommendations, both sets of principles placed a strong emphasis on accurate time-series pattern representation, data transparency, and unambiguous labeling. In line with Stephen's Principle 7 and the recommended usage of grouped bar charts and small multiples, both sets emphasized the significance of organizing comparisons inside a single view (e.g., presenting all specializations or items side-by-side). Both lists also acknowledge how crucial it is to keep color schemes constant in order to facilitate user recognition and lessen cognitive burden.

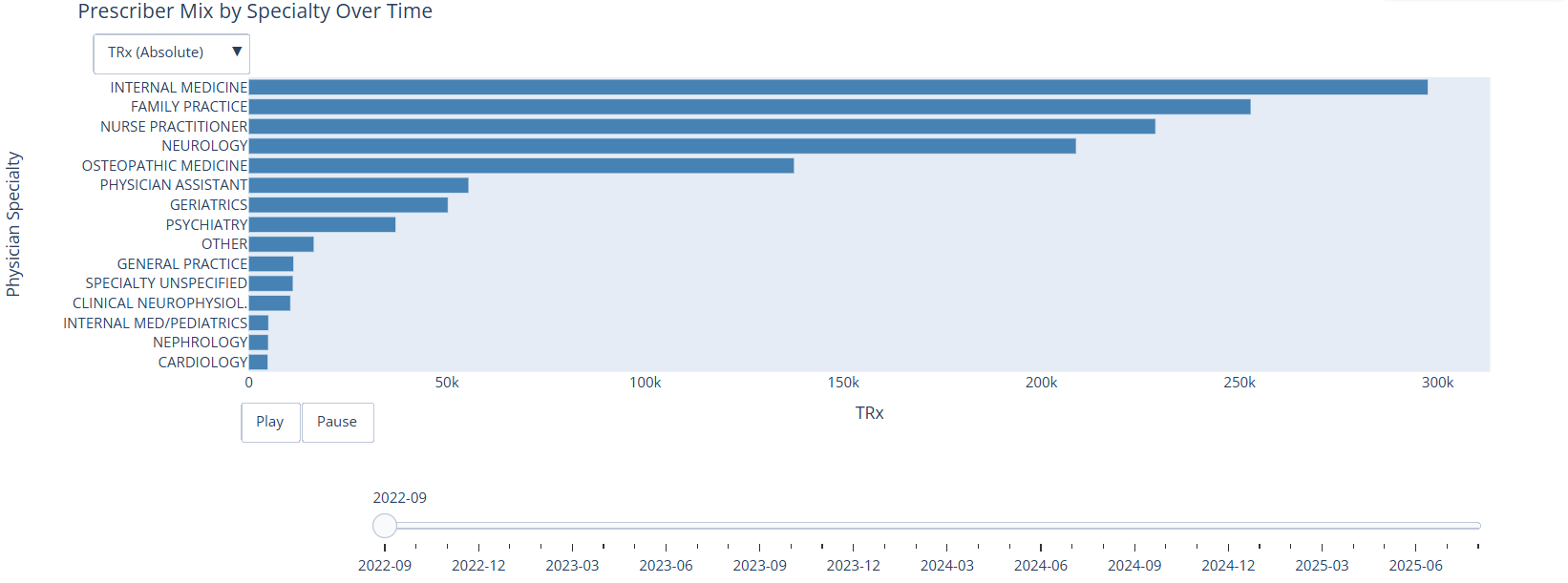
* 1. **Final Dashboard**

I created 4 plots:

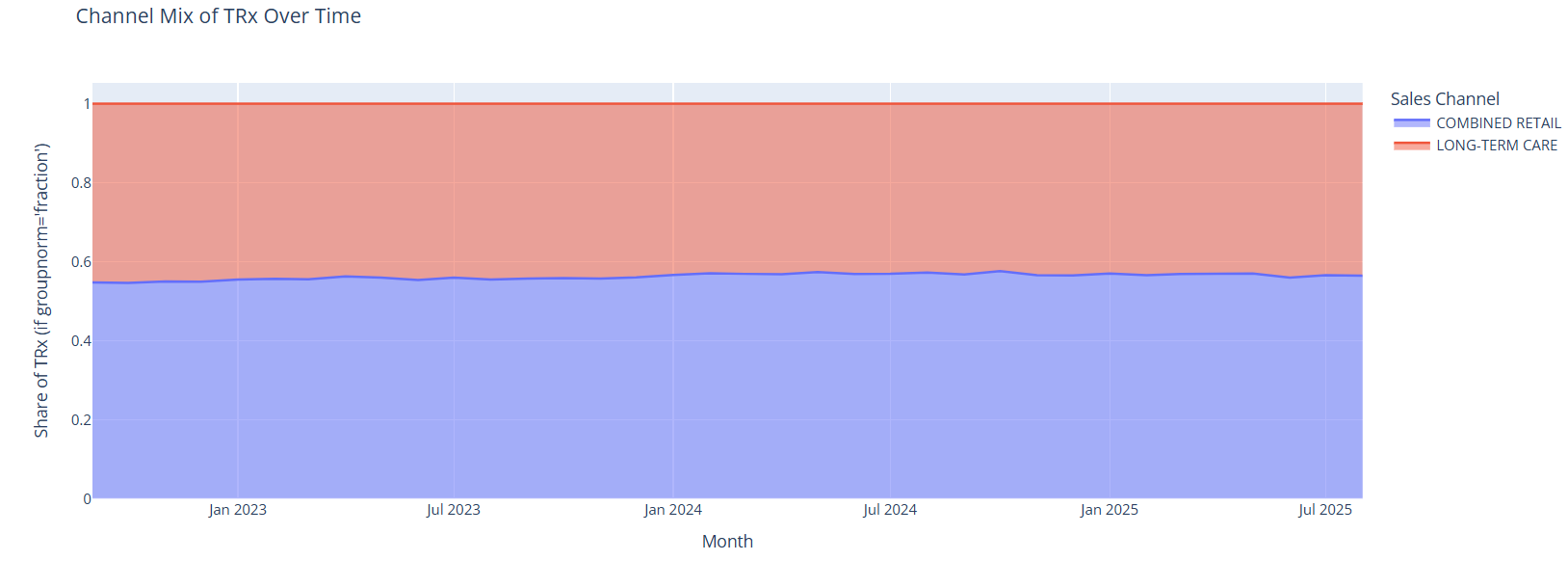
**Total TRx Over Time by Product**

****

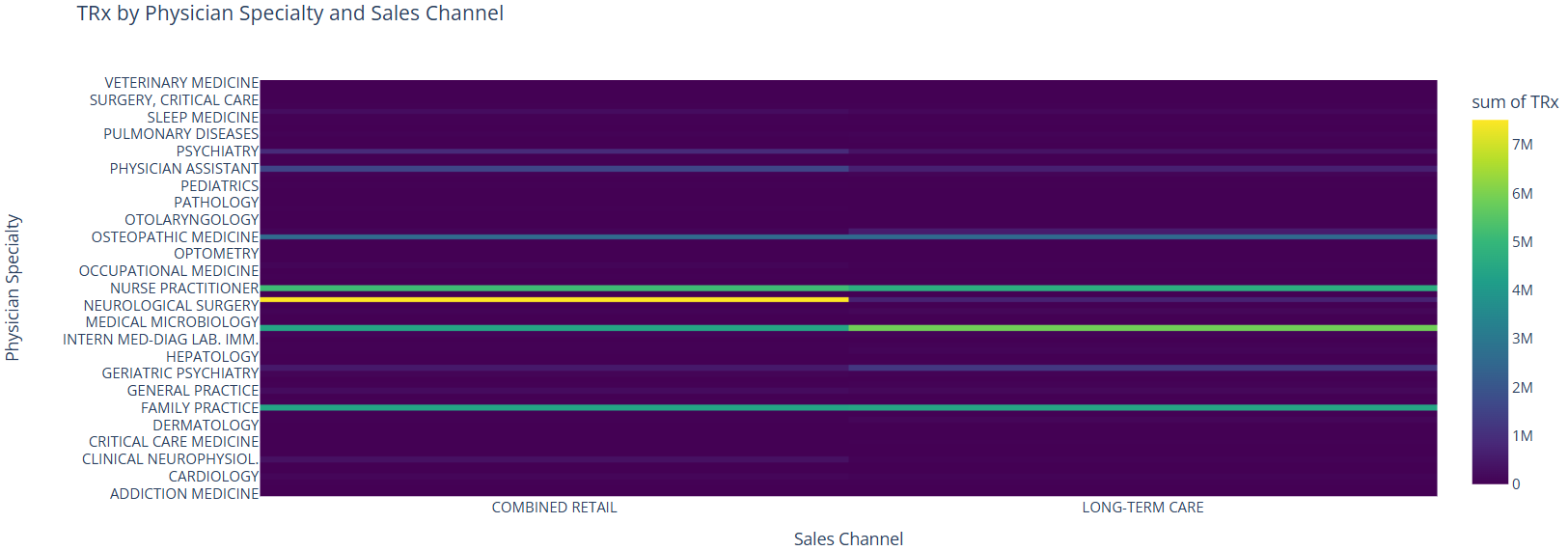
**Prescriber Mix by Specialty Over Time**

****

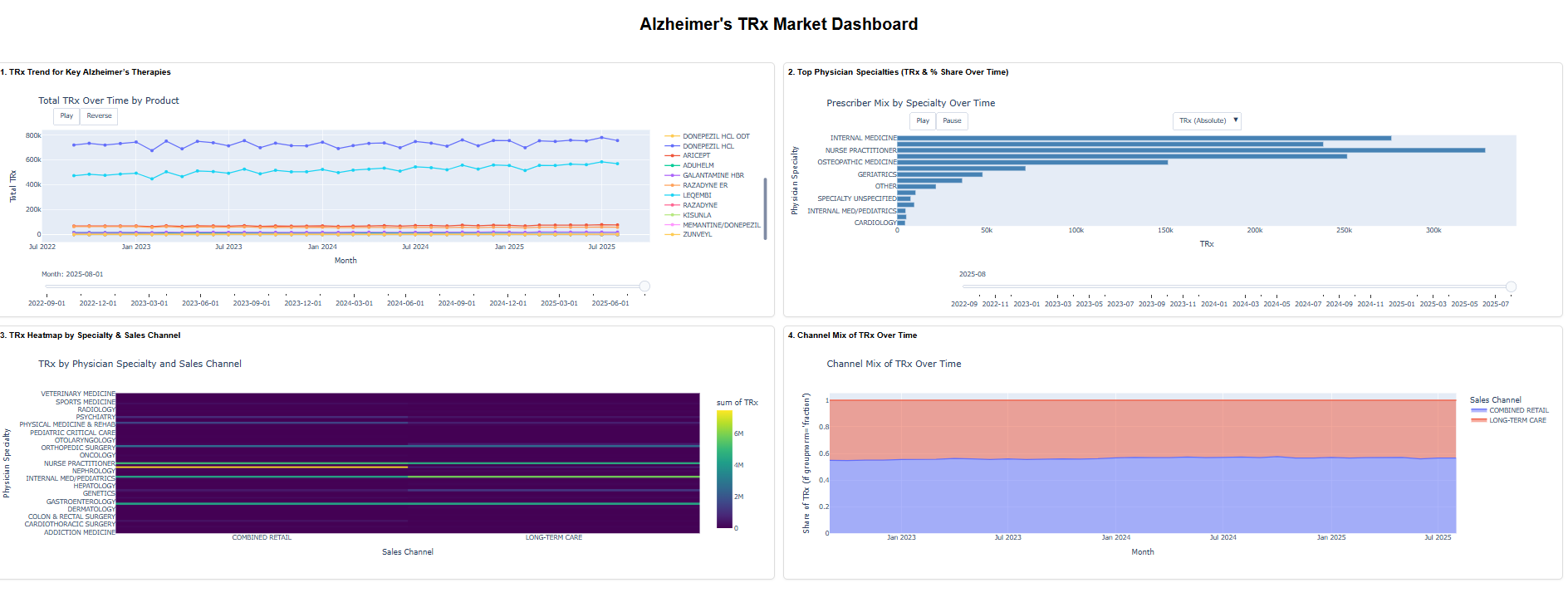
**Channel Mix of TRx Over Time**

****

**TRx by Physician Specialty and Sales Channel**

****

**Then I combined all these 4 plots into a 2X2 grid HTML file**



**Key Interesting Findings from the Dashboard**

* While newly branded medications have delayed but rising acceptance patterns, established Alzheimer's therapies account for the majority of TRx volume. Generics like donepezil, memantine, and rivastigmine routinely top the market.
* Although Adlarity and Leqembi show clear early growth patterns, their TRx levels are still significantly lower than those of traditional medicines, which is a reflection of both the time of market debut and the smaller prescriber uptake.
* Internal medicine and nurse practitioners also make significant contributions, although neurology and primary care continue to be the two major specialty drivers, accounting for the largest share of TRx throughout the dataset.
* Specialty TRx shares have been comparatively consistent over time, although slight changes show that non-neurology specialties are becoming more involved, which may signal that general practice settings are becoming more comfortable with Alzheimer's treatment.
* The bulk of prescriptions are filled through retail channels, however the dashboard shows a small but steady increase in mail-order channels, which may be a sign of shifting patient preferences or payer-driven dispensing practices.
* Primary care in retail settings provides the largest TRx, according to Specialty × Channel research, demonstrating the continued importance of broad-access retail dispensing in Alzheimer's therapy.
* Strong TRx performance in neurology across retail and non-retail channels suggests that neurologists handle more complicated patient profiles that might need institutional support or long-term care.
* The constant yet concentrated non-retail channels (such long-term care) indicate consistent, predictable TRx activity in institutional patient groups.
* Changes in prescription fulfillment practices, the expansion of telemedicine, or insurer channel steering may all be contributing factors to the evolution of the channel mix, which shows a slow transition from strictly retail to more varied channel distribution.
* Together, TRx trends, specialty contributions, and channel behavior across all visualizations show unrealized potential, especially in neglected specialty–channel intersections that can be the focus of future engagement or advertising.  
  1. **Conclusion:**

This dashboard project converts complicated Alzheimer's prescription data into understandable, useful information that immediately aid in making strategic decisions. The dashboard provides precise, understandable, and significant depictions of TRx trends, specialist behaviors, and channel success by coordinating the visual design with accepted ideas from Tufte, Gestalt theory, and Stephen Few. A cohesive, insight-driven analytical tool was produced by specifically designing each visualization to address one of the four main analytical queries.

The dashboard's key findings show which medicines are gaining or losing market momentum, which physician specialties are driving growth, how TRx volume is shaped by specialty–channel intersections, and how the channel mix as a whole is changing over time. These insights show possible avenues for focused involvement as well as present performance trends. Users may examine the data at both a high-level and a detailed level without experiencing cognitive overload because to the incorporation of interactive filters and well-organized images.

All things considered, this project shows the value of combining solid analytical underpinnings with careful visual design. The final dashboard serves as a scalable, repeatable paradigm for next market analytics and decision-support tools in addition to successfully addressing the business objectives.

**APPENDIX:**

**Html:**



**DataSet:**



**Notebook:**

