Operations Analysis:

Performing analysis on data generated by ChatGPT

Emmanuel Sanchez

MGMT 339

Spring 2025

Fridays – 1pm

May 5, 2025

This project involves a fictitious dataset that is generated by ChatGPT. Any similarities in data are purely coincidental. With the emergence of Artificial Intelligence (AI) there are plenty of generative chatbots based on Large Language Models (LLM) that are generated collaboratively by Data Engineers, Data Scientists and Data Analysts. It is common practice for Data Analysts to ask AI chatbots to generate randomized datasets to be analyzed to practice analysis of big data. While there are plenty of sources on the internet that can be analyzed, the quality of the data as well as the accessibility may not be suitable for consumption by a data analyst. In a prompt with ChatGPT, I asked for a fictious dataset by a company so I can analyze their operations. The response I got was:

"*Let's assume this company is called* ***"BrightSupply Co."****, and it's a mid-sized e-commerce company specializing in office supplies."* My only error after it generated data regarding the operations of this company was that it didn't generate any million row datasets. Typically, in big data programs would have to be used to analyze insights and trends due to the high amount of data. In this case there was 5 datasets generated and a maximum of 150 rows in the largest dataset. While I could generate some key insights quickly, it's always best practice to generate insights via analysis tools.

At the time that this report is written which was Spring semester 2025, TitanGPT which is a free version of ChatGPT specialized for California State Fullerton was unveiled. I believe one of the many misunderstandings with AI and chatbots are that they know everything. While I am trying not to digress from the report to be generated this is an import facet of AI that must be understood: AI will only be able to understand data that is accessible. TitanGPT was generated specifically to answer questions regarding CSUF and their operations whether it is how to add or

drop classes and when deadlines are. When I attempted to task it with generating a randomized dataset along with analysis for this report, it was unable to do so.

With any data analysis, we are looking to answer any questions. With *Bright Supply Co.,* we want to know about their operation and with the information given we want to answer the question: Which warehouse has the best performance and why? As well as can anything be done? Looking at the dataset, we are given five tables listed in no particular order: “Operations Data”, “Inbound Shipments”, “Inventory Transactions”, “Outbound Shipments”, “Suppliers Data”. All the data in each table linked by a “primary key” which is the “Warehouse ID”. The only outlying table that has a relationship to all warehouses but not transactions which is the Supplier which stocks each warehouse individually. We will look at each table to answer the business question being asked. The way all the tables are structured and related gives us the following star schema:

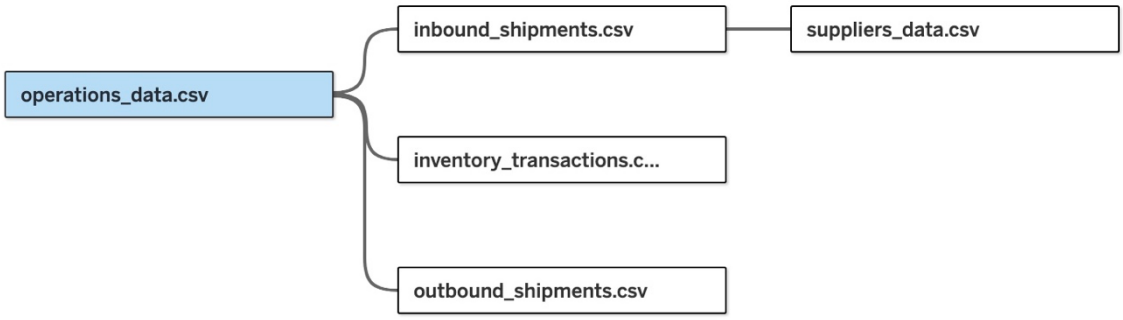


Figure 1: Star Schema Chart

To start, we need to assess which warehouse has the best performance by creating a KPI for the warehouse under the table “operations data”. This table has columns such as Warehouse ID, warehouse location, total orders processed, average processing time, inventory level, back orders, fulfillment rate, labor hours and operational cost which are specific to the warehouse and

the month of operations. Running calculations, I created new categories: cost and orders processed per hours and used those two factors along with fulfillment rate to create a KPI derived from the following formula:

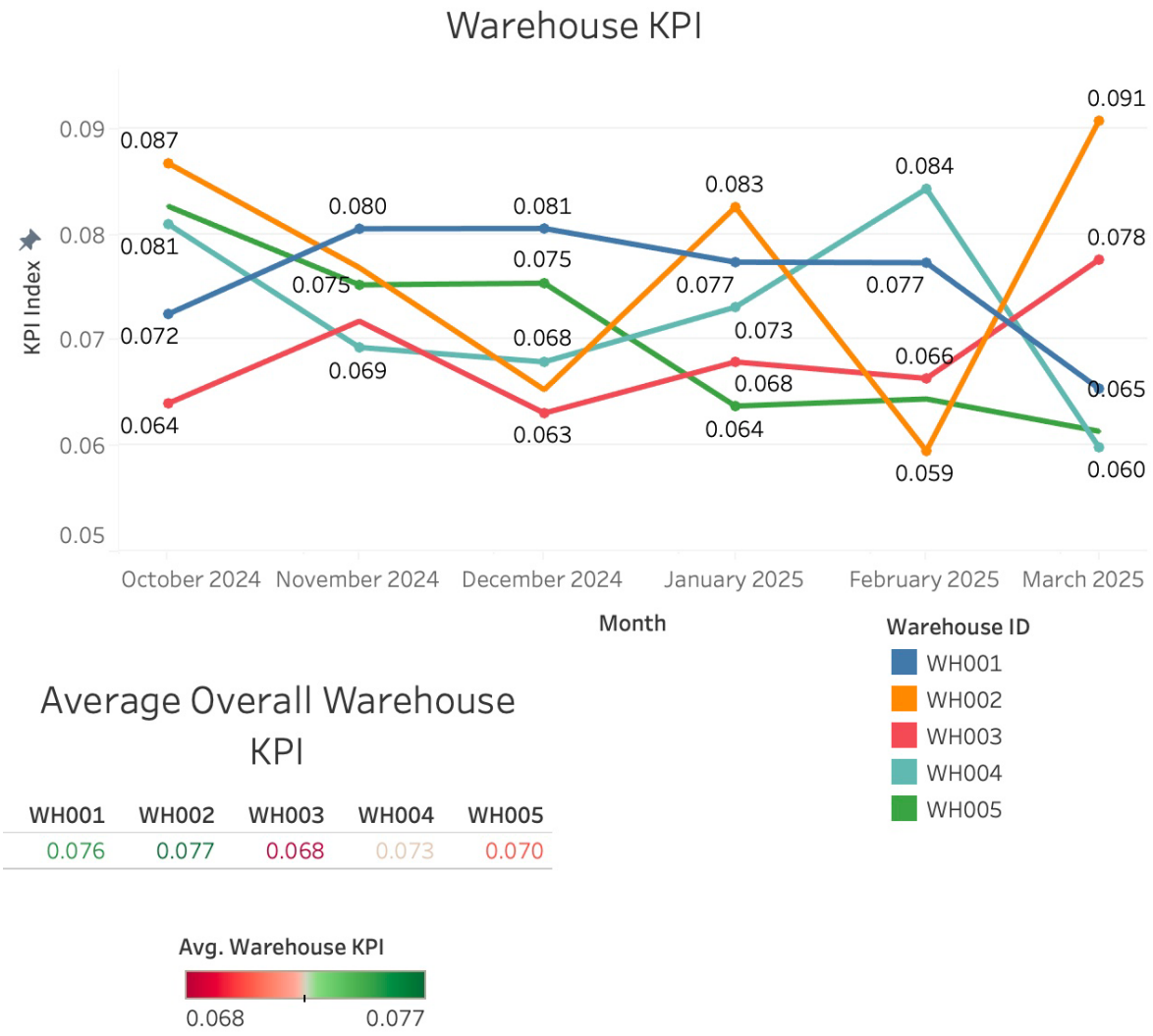
Equation 1Warehouse Efficiency KPI Formula

Warehouse Efficiency KPI

Formula:

$$\text{Warehouse Efficiency KPI} = \frac{\text{Orders Processed per Hour} \times \text{Fulfillment Rate}}{\text{Cost per Warehouse per Hour}}$$

Charting the respective numbers give us this dashboard:



Graph 1Warehouse Efficiency KPI

According to these numbers it looks like Warehouse 002 (WH002) is the best performing warehouse throughout the past 6 months. We can take a look at the relevant data to see what factors may have contributed to the performance of this warehouse.

The table with only relevance to the warehouses is the supplier table. There are five suppliers which each supplies a difference type of product to each warehouse. The categories of importance that they provide are Lead Time, On Time Delivery Rate and Average Shipment Quality. In order to answer the business question regarding the supplier's performance was

creating a Key Performance Index(KPI) factoring the three previously mentioned categories. The KPI created was the Shipment Efficiency Score(SES) which is a number where the higher the number, the better the score. The formula is as follows:

Equation 2 Shipping Efficiency Score

Formula:

$$\text{SES} = \left(\frac{\text{Avg Quantity of Items Shipped}}{\text{Max Quantity}}\right) \times (\text{Fulfillment Rate}) \times \left(\frac{\text{Max Lead Time} - \text{Lead Time in Days}}{\text{Max Lead Time}}\right)$$

The formula generates the following chart:

Table 1Supplier Data

Supplier Data KPI

	SUP001	SUP002	SUP003	SUP004	SUP005
Avg Shipment Quantity	5,974	5,277	6,010	6,996	5,431
Lead Time Days	11	6	10	15	10
On Time Delivery Rate	88%	91%	88%	89%	89%
Supplier KPI	8.01	15.93	10.07	0.00	9.11

Supplier Damage Rate

	SUP001	SUP002	SUP003	SUP004	SUP005
Damaged Units	101	99	58	81	86
Quantity Received	25,395	29,387	27,823	23,844	37,514
Damage Rate Overall	0.40%	0.34%	0.21%	0.34%	0.23%

The chart indicates that based on the Shipping Efficiency Score, Supplier SUP002 has the best performance relevant to the determining factors in the KPI. Furthermore, we can take look at the warehouses and suppliers to cross reference their main associations because although there are 5 suppliers and 5 warehouses, they don't all do business with each other.

Table 2 Supplier -Warehouse Relationship

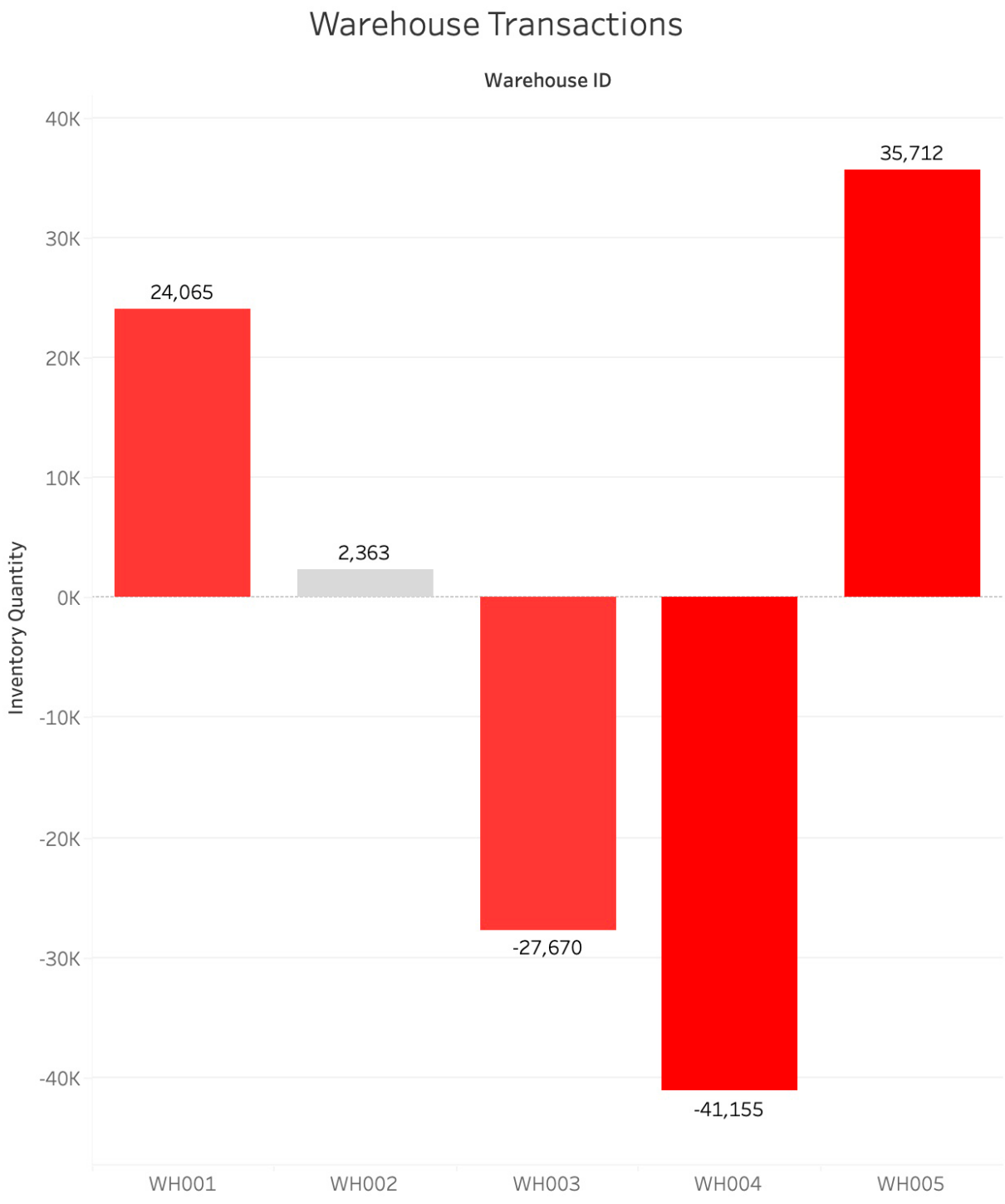
Supplier - Warehouse Cross Reference

	Supplier ID				
	SUP001	SUP002	SUP003	SUP004	SUP005
WH001	2	1	2	2	5
WH002	1	3	1	1	
WH003	1		1	2	1
WH004		1		1	
WH005	2	1	2		

From the past two charts, we can determine that Supplier 002 does have best SES and also subsequently does the most business with warehouse 002. Also looking at Table 1, the supplier damage rate in the shipped goods is nominal to the performance factor that they were not deemed significant enough to alter any scores.

Next, we will look at the “inventory transactions” table to see how each warehouse is managing its inventory. There are listed monthly transactions of everything that has been shipped into stock inventory and outbound shipments to customers. The following graph will

show the average for those 6 months.



Graph 2 Warehouse Inventory Levels

The graph shows that Warehouse 002 has inventory closest to zero indicating that it never has too much surplus nor possibly backfilled on orders going out to customers. Looking at the numbers, with Warehouse 004 leading the way in negative inventory and Warehouse 003 not far behind, checking their respective backorder numbers, Warehouse 004 had 1,829 while Warehouse 003 had 1,906 which checks out in regard to their negative inventory.

Lastly, we will take a look at what seems to be an irrelevant contribution to warehouse performance which are the outbound shipments. Something to consider is that once it's in the pipeline to the customer, it is out of the companies' hands. Four various shippers amongst DHL, UPS, UPHS and FedEx are the responsible companies for getting the shipments to the retail outlets.

Table 3 Outbound Deliveries

Overall On Time Deliveries

WH001	WH002	WH003	WH004	WH005
13,950	13,845	17,867	14,799	15,189
19,699	20,940	24,102	22,141	20,735
70.82%	66.12%	74.13%	66.84%	73.25%

Table 3 shows Warehouse 002, our best performing warehouse having the worst percentage of on time deliveries as well the amount delivered on the lower end, but again these numbers may be independent of the warehouse due to it being the responsibility of the shipping company.

To conclude this report, we should note a few things. ChatGPT was used to generate the KPI formula for a non-biased index. The only problem where a bias can come in is what the

human prompted AI to do. It was told to look at three factors into a formula so the question arises, should it also have considered on time shipments as well as the number of items delivered? The consistency seems to come from what company supplied the supplies, and based off of one table, Supplier 002 seemed to be the best performing due to the factor of the lead time for products. The lead time of 6 days compared to the next highest which was 10 days could make the difference between one week or two weeks when having a stable inventory as well as having fewer back orders.