



Amazon E-Commerce

Aidan, Saurabh, Olivia,
Boran, Vivek, Kshitij

Table of contents

01

Topic/Company

02

**Data Definition
Language**

03

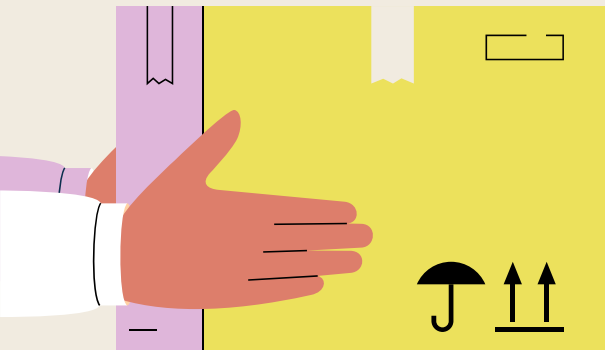
**Data Warehouse
EDL**

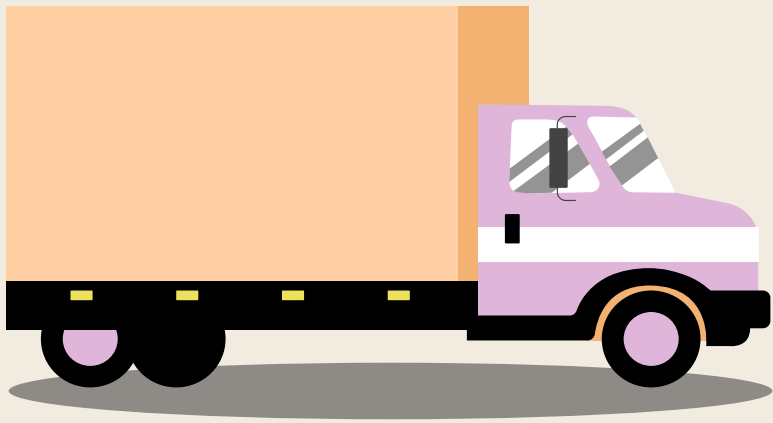
04

**Data Lake Model &
Analytics**

05

Reflection



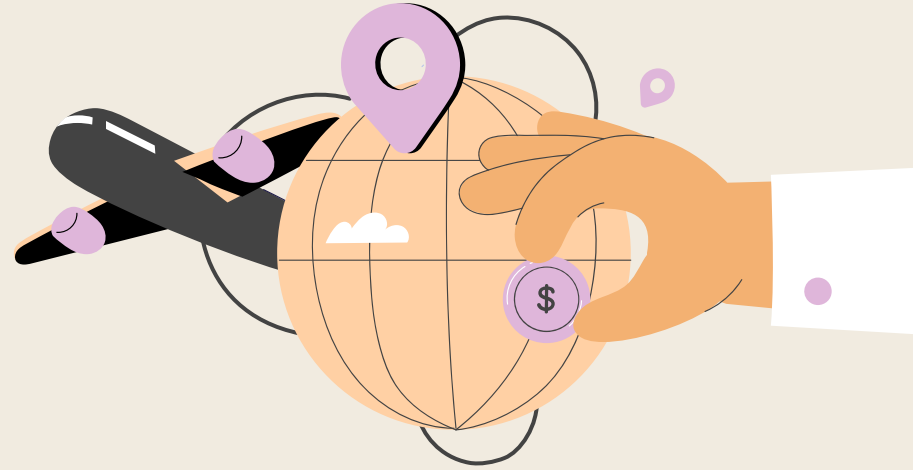


01

Topic / Company

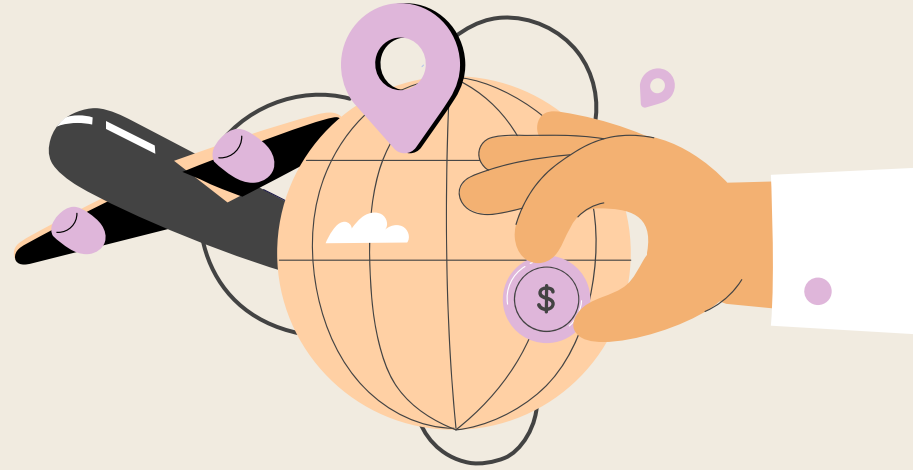
Why Amazon?

- Largest e-commerce enterprise.
- Has over 300 million active customer accounts and over 1.9 million selling partners worldwide.
- Launched numerous private-labels such as Amazon Basics, Amazon Fresh, Amazon Essentials, Prime, Solimo, Kindle, AWS and many more.
- Ecosystem built around constant innovation and change.
- To comprehend the organizational structure and different components of the business



Data Strategy

- Amazon primarily uses Offensive Strategy.
- Amazon targets competitors' keywords so that if a customer searches for the competitor they will be served an ad from you on the search results page.
- Has an aggressive approach of expansion and innovation.
- Build and architecture to manage data volume and velocity

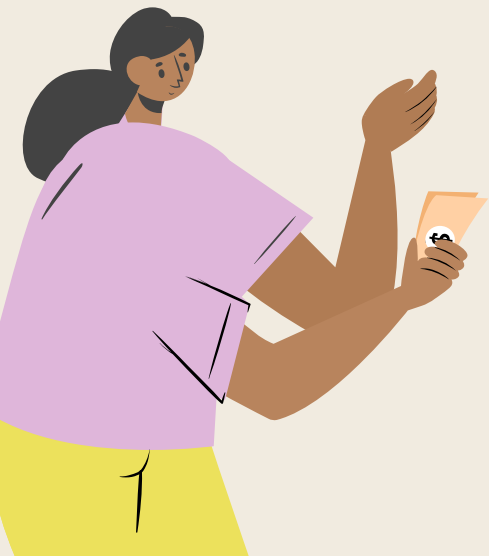




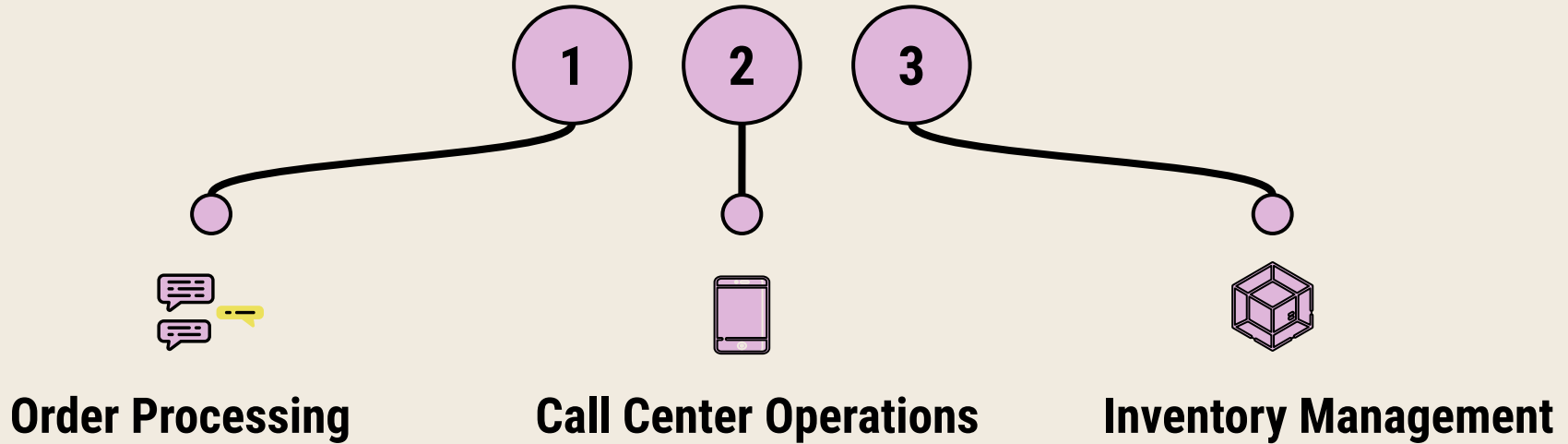
02

DDL

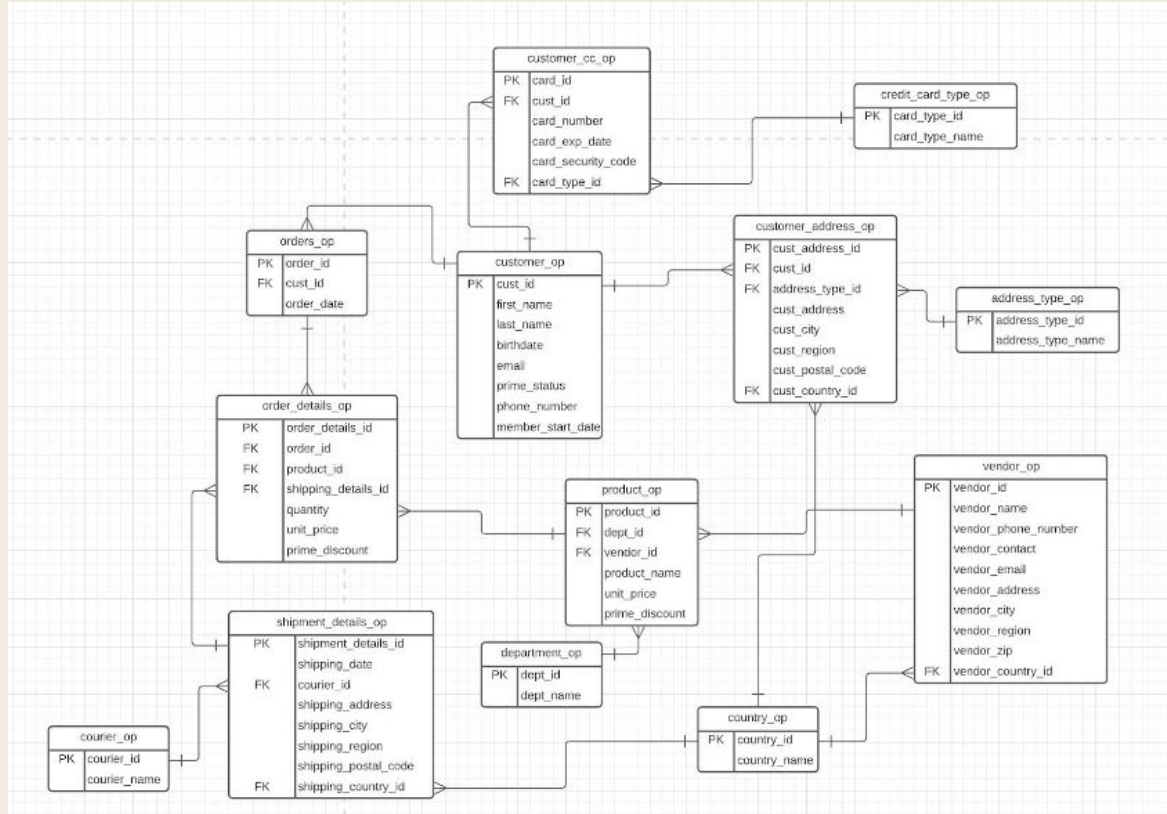
Data Definition Language



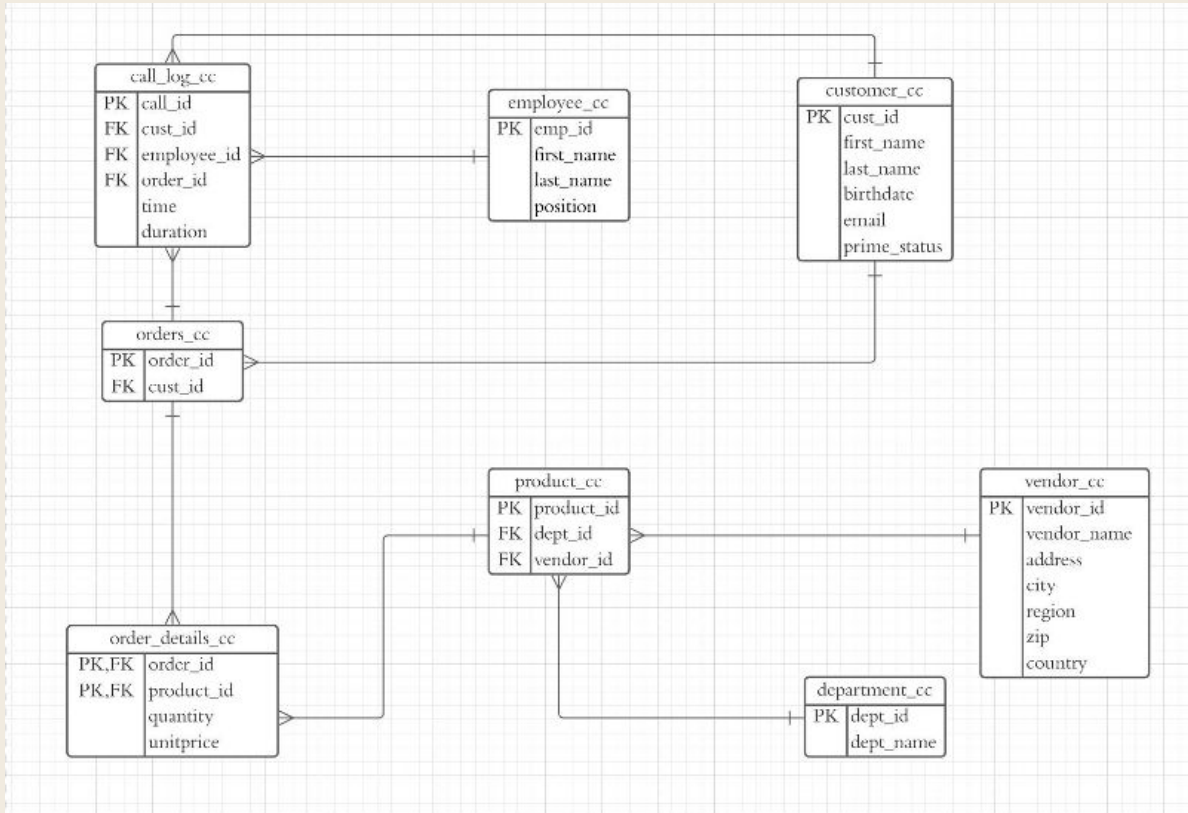
Three Types of Transactions



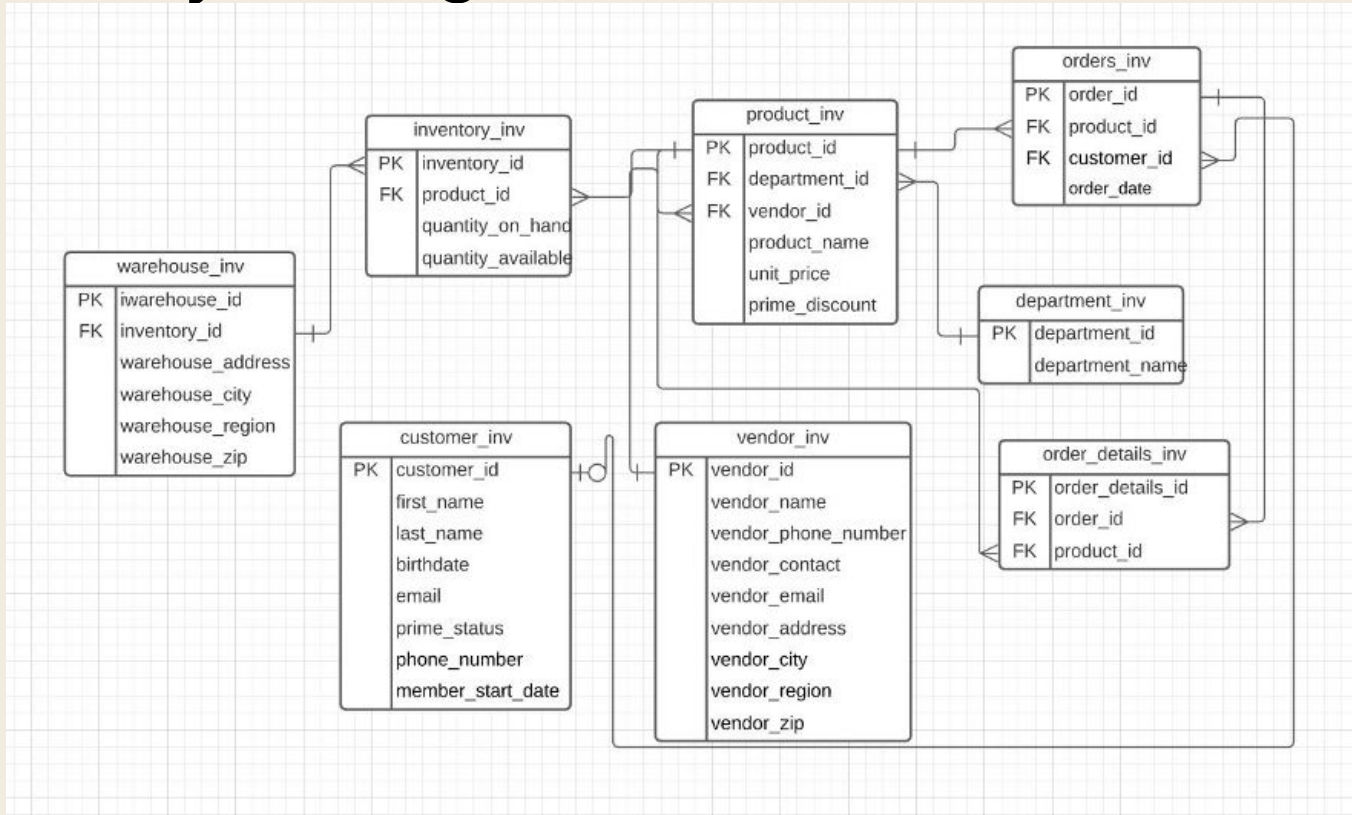
Order Processing



Call Center Operations



Inventory Management



Queries from Transaction Tables

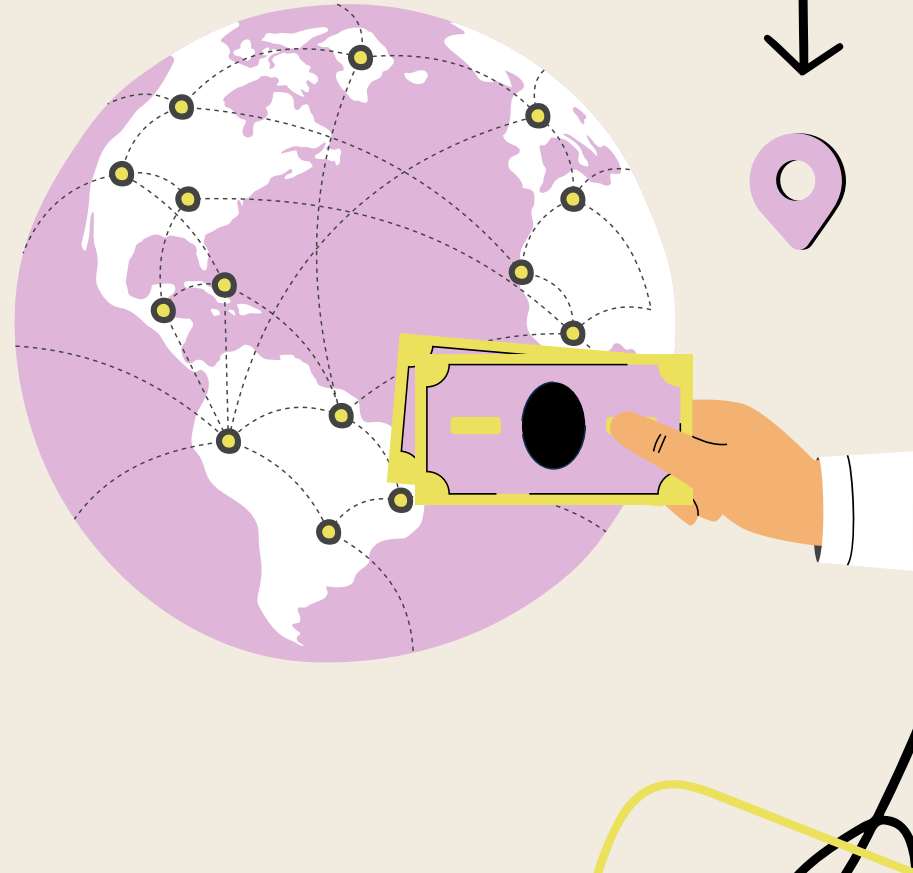
	Shipping Vendor	# Late Deliveries
1	USPS	62
2	UPS	12

Evaluating our shipping vendors by seeing how many of their deliveries aren't on time

	Product	# Complaints	5-Star Rating
1	YETI Rambler 20 oz.	11	3.85
2	Keurig K-Mini Coffee Maker	9	2.41
3	Hydro Flask Standard Mouth Bottle with Flex Cap	7	1.54
4	Fire TV Stick with Alexa Voice Remote	7	1.02
5	Magic Bullet Blender	6	0.63


Looking into the most complained about products on customer service calls vs. online rating

03 Data Warehouse & ETL

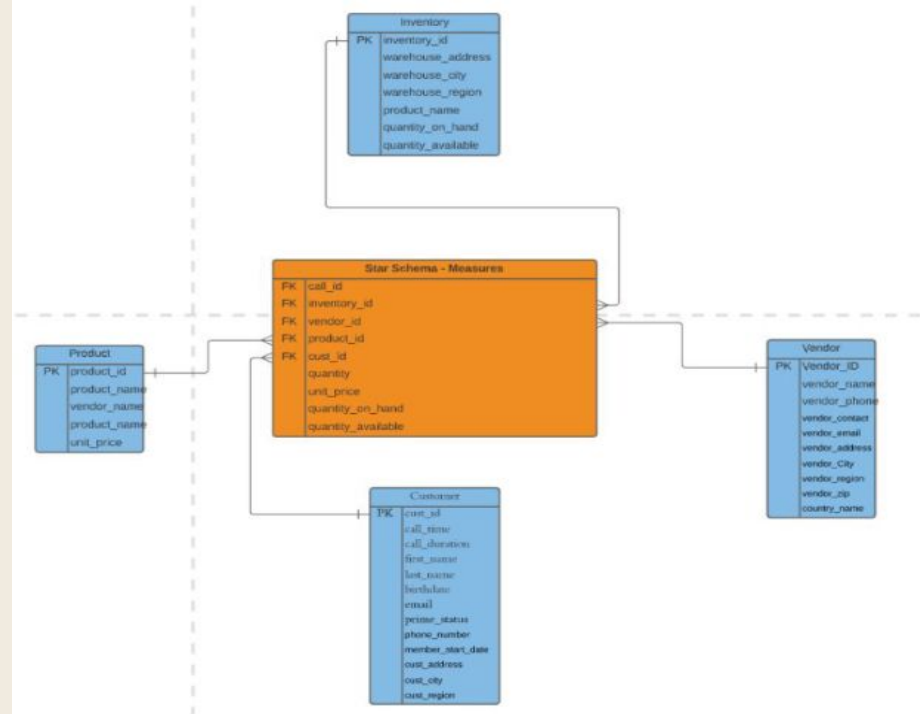
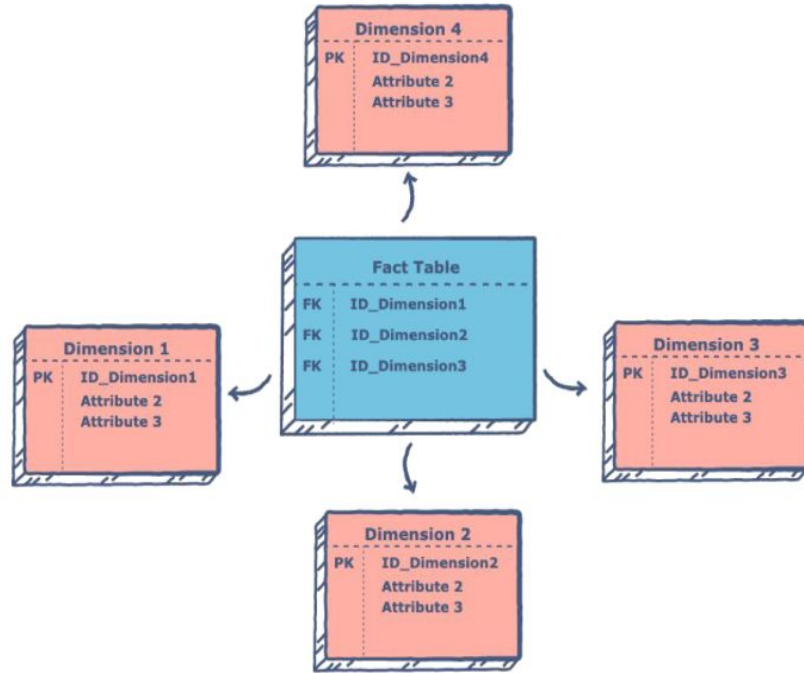




Design Decisions

- Star Schema - Product, Vendor, Inventory, Customer Dimensions
 - Analyze call center call volume, warehouse inventory, location, etc.
 - Identified SCDs; reduced data volume
 - Essential in mapping trends to better understand transactional operations
 - ETL vs ELT
- 

Data Warehouse

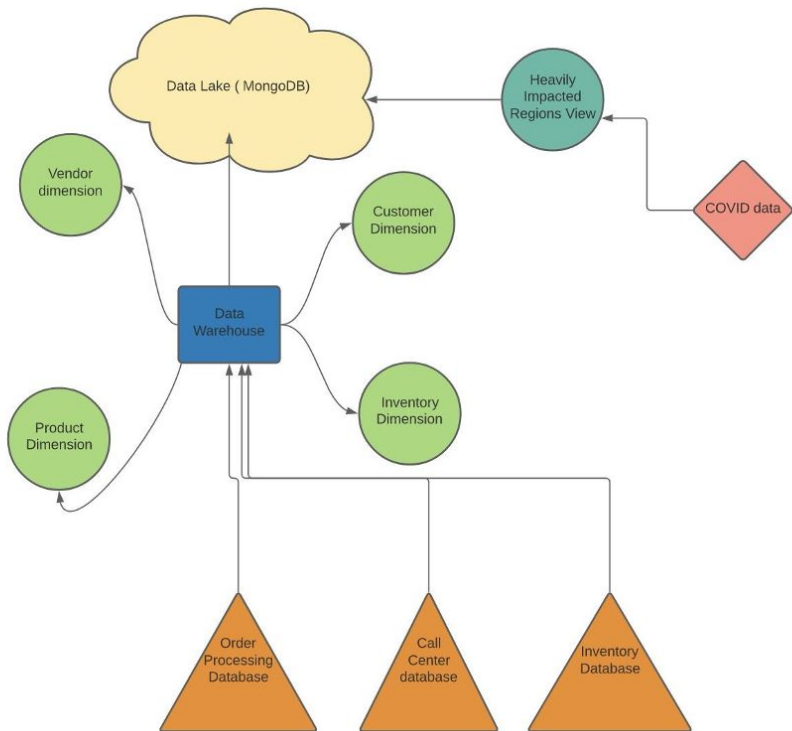


04

Data Lake Model & Analytics

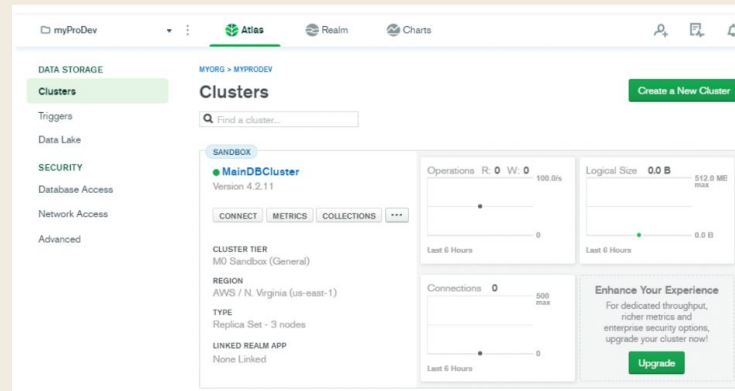


Data Lake Model-MongoDB and AWS S3



A screenshot of the Amazon S3 console showing the "improjeckshittij" bucket. The "Objects (17)" list is displayed with columns for Name, Type, Last modified, Size, and Storage class.

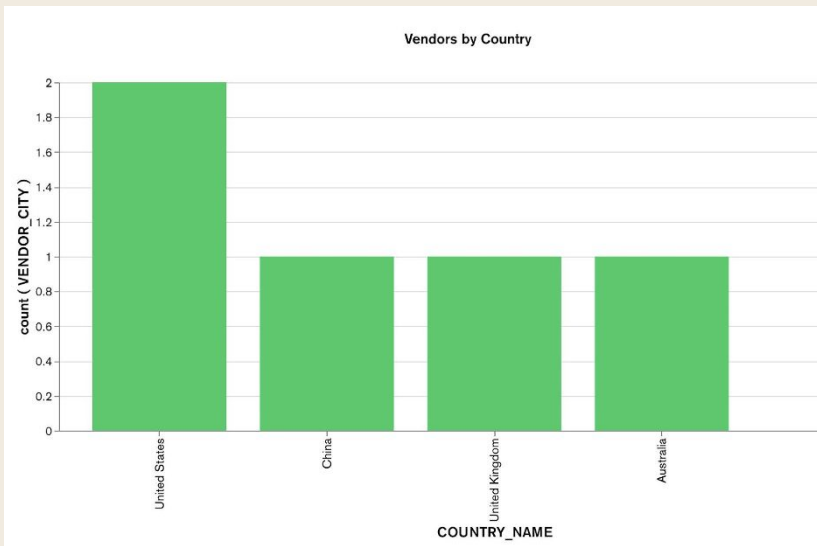
Name	Type	Last modified	Size	Storage class
Call_Center.csv	csv	December 1, 2022, 23:54:44 (UTC-06:00)	233.0 B	Standard
Call.csv	csv	December 1, 2022, 23:54:44 (UTC-06:00)	12.0 KB	Standard
CUSTOMER_DIM_DATA_TABLE.json	json	December 1, 2022, 22:19:50 (UTC-06:00)	2.1 KB	Standard
Customer.csv	csv	December 1, 2022, 23:54:45 (UTC-06:00)	23.4 KB	Standard
Employee.csv	csv	December 1, 2022, 23:54:45 (UTC-06:00)	1.6 KB	Standard
HEADQUARTES_DIM_DATA_TABLE.json	json	December 1, 2022, 22:19:50 (UTC-06:00)	1.2 KB	Standard
Order_Product.csv	csv	December 1, 2022, 23:54:45 (UTC-06:00)	12.4 KB	Standard
Order.csv	csv	December 1, 2022, 23:54:46 (UTC-06:00)	43.3 KB	Standard
Product.csv	csv	December 1, 2022, 23:54:46 (UTC-06:00)	114.0 B	Standard
PRODUCT_DIM_DATA_TABLE.json	json	December 1, 2022, 22:19:50 (UTC-06:00)	854.0 B	Standard
Product_Supplier.csv	csv	December 1, 2022, 23:54:47 (UTC-06:00)	21.2 KB	Standard
Product_Warehouse.csv	csv	December 1, 2022, 23:54:47 (UTC-06:00)	2.1 KB	Standard
Shipping_Vendor.csv	csv	December 1, 2022, 23:54:48 (UTC-06:00)	356.0 B	Standard
Supplier.csv	csv	December 1, 2022, 23:54:48 (UTC-06:00)	11.9 KB	Standard
VENDOR_DIM_DATA_TABLE.json	json	December 1, 2022, 22:19:51 (UTC-06:00)	461.0 B	Standard
Warehouse.csv	csv	December 1, 2022, 23:54:48 (UTC-06:00)	394.0 B	Standard



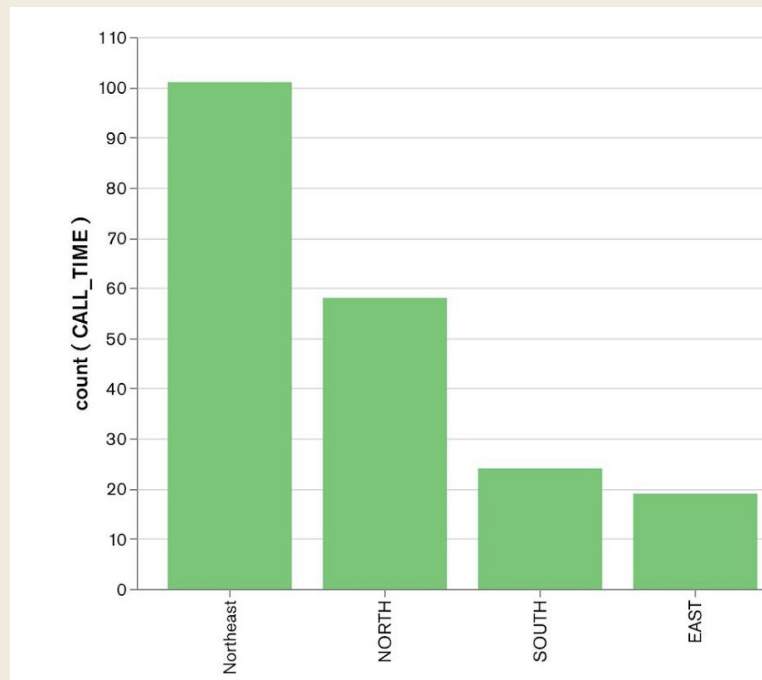
Analytics- MongoDB Data Lake

- MongoDB charts

Vendors count by country

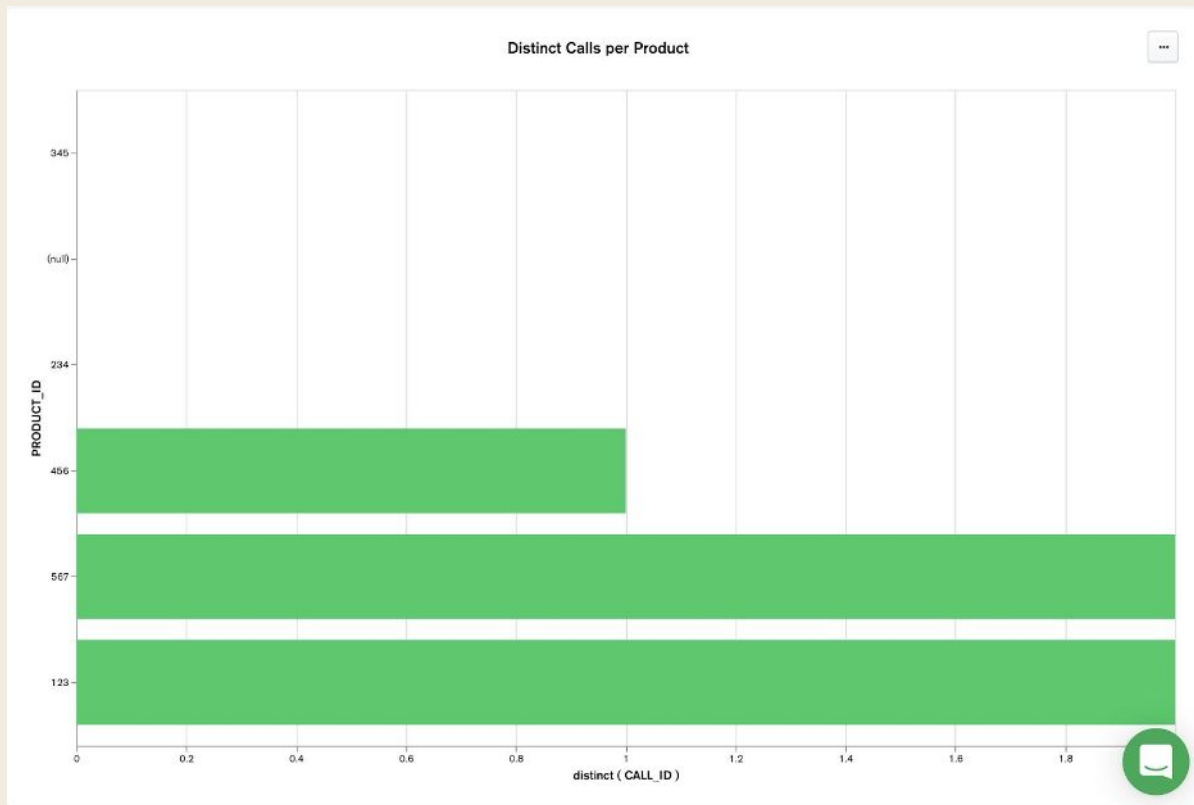


Customer region distribution



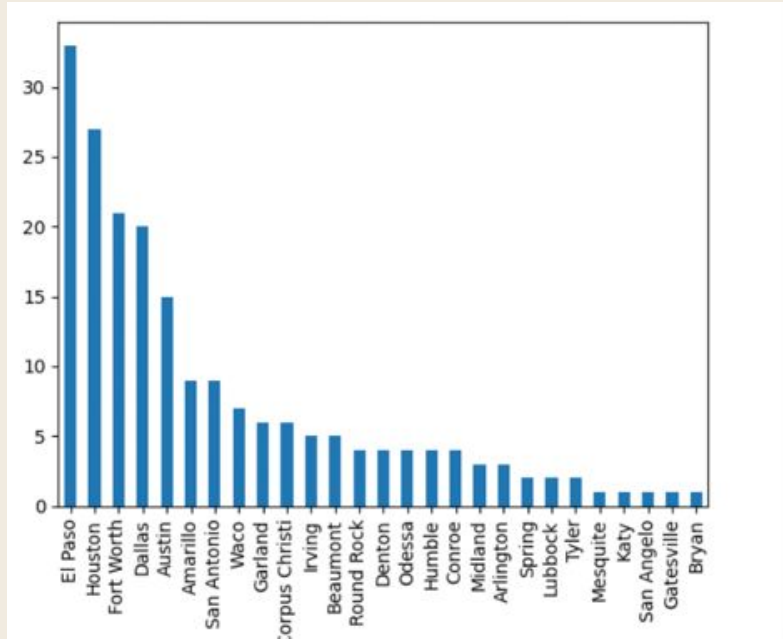
MongoDB con't

Average duration call per customer

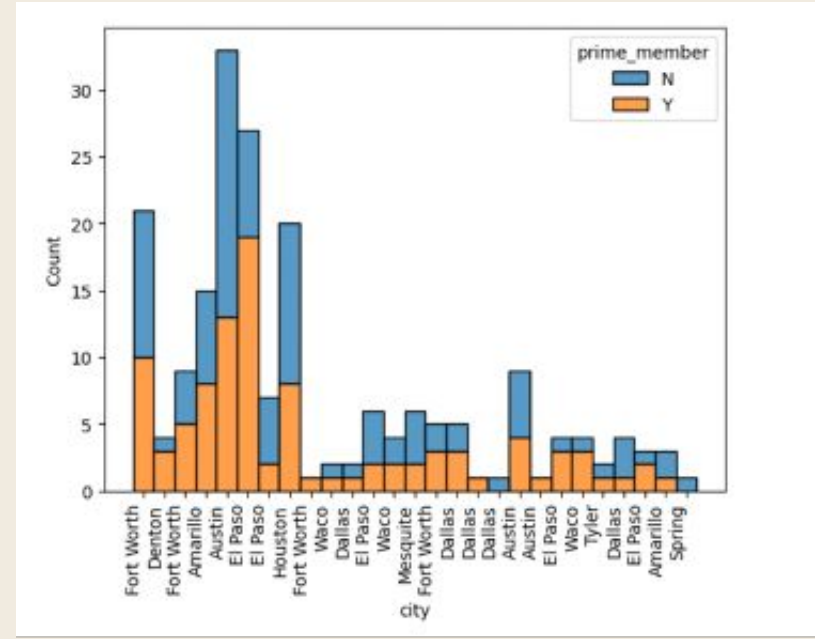


Analytics - Amazon S3 Data Lake

Number of customers per city

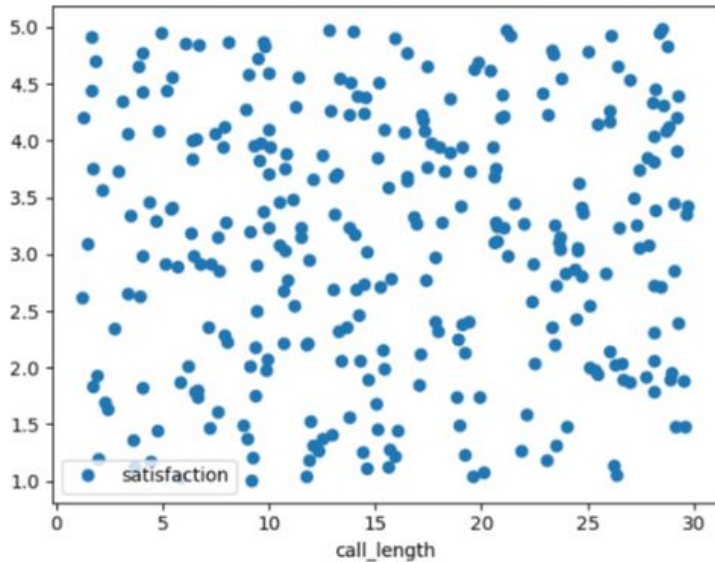


No. of Prime members per city

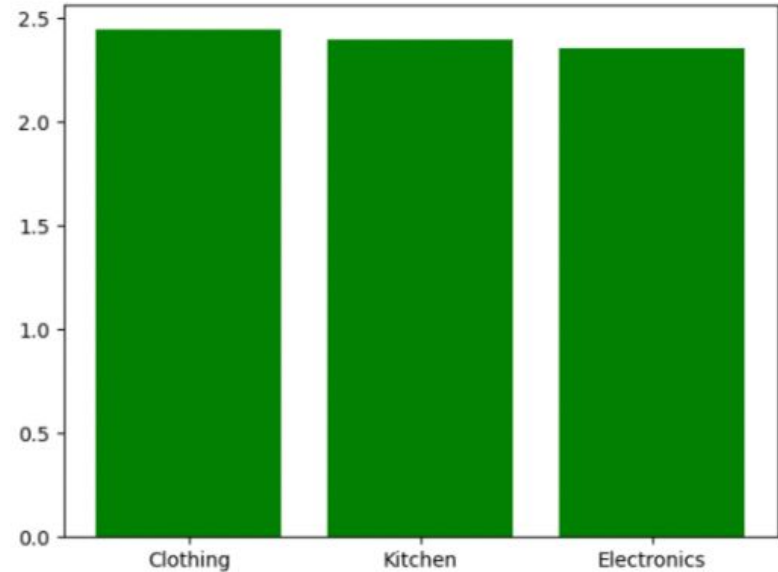


Amazon S3 Data Lake continued..

Relation between call length
and customer satisfaction

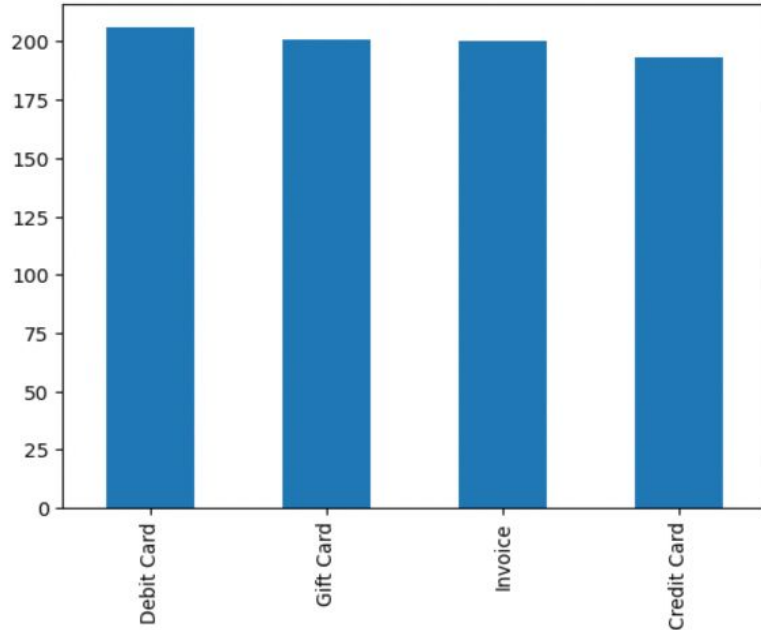


Average rating for
each product category

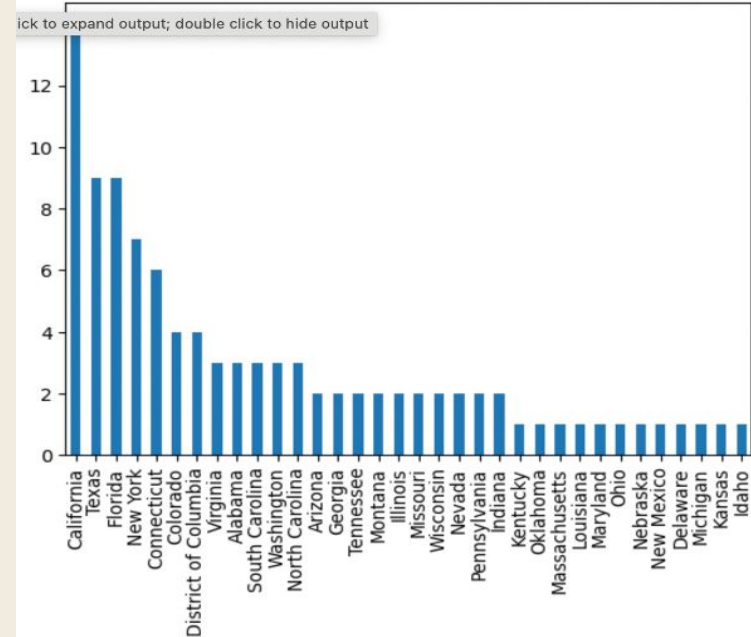


Amazon S3 Data Lake continued..

Number of customers using each payment method

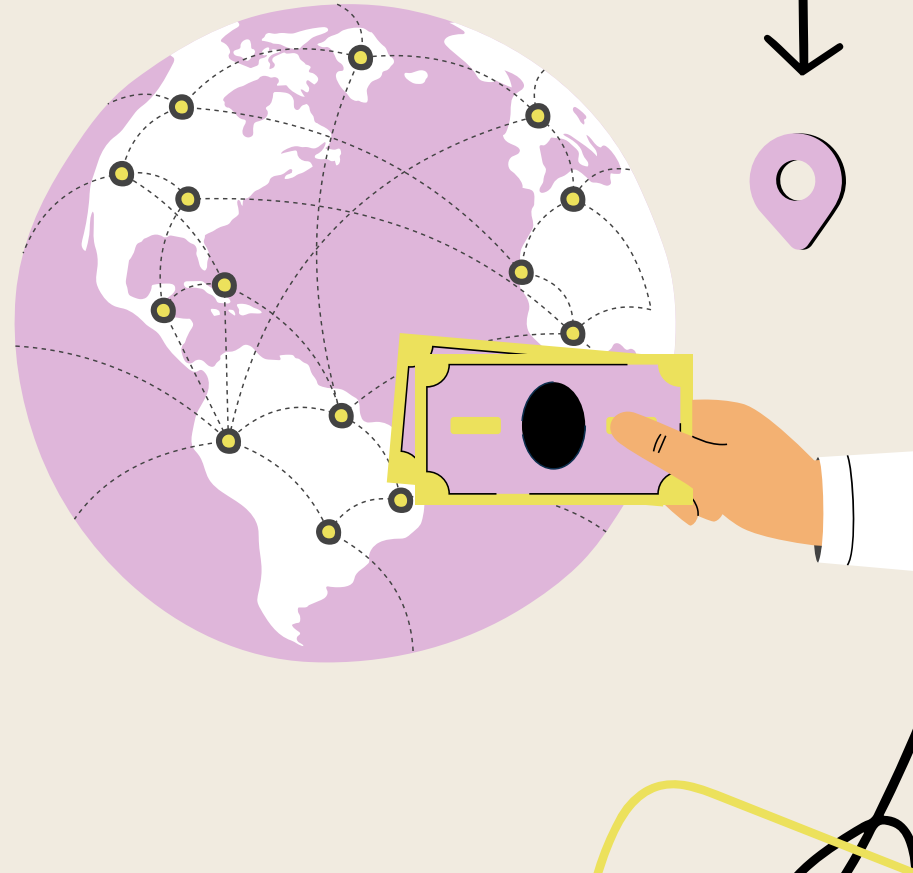


Number of suppliers per state



05

Reflection

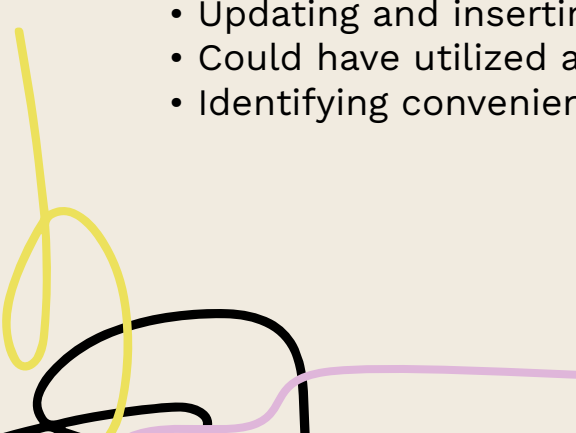


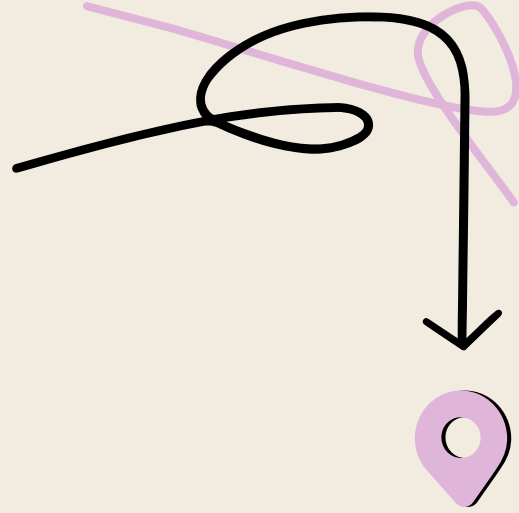


Opportunities :

- Times series data
- Additional work with data lakes
- Emphasize dealing with big data
- Streaming data

Changes:

- Access to more data for various departments
 - Updating and inserting the data was difficult - identify a more efficient method
 - Could have utilized an alternate platform to create a data lake
 - Identifying convenient ways to seed realistic data
- 



Thanks!

Do you have any questions?

