theLook project

Tawei Lo

Content

Insights from the data
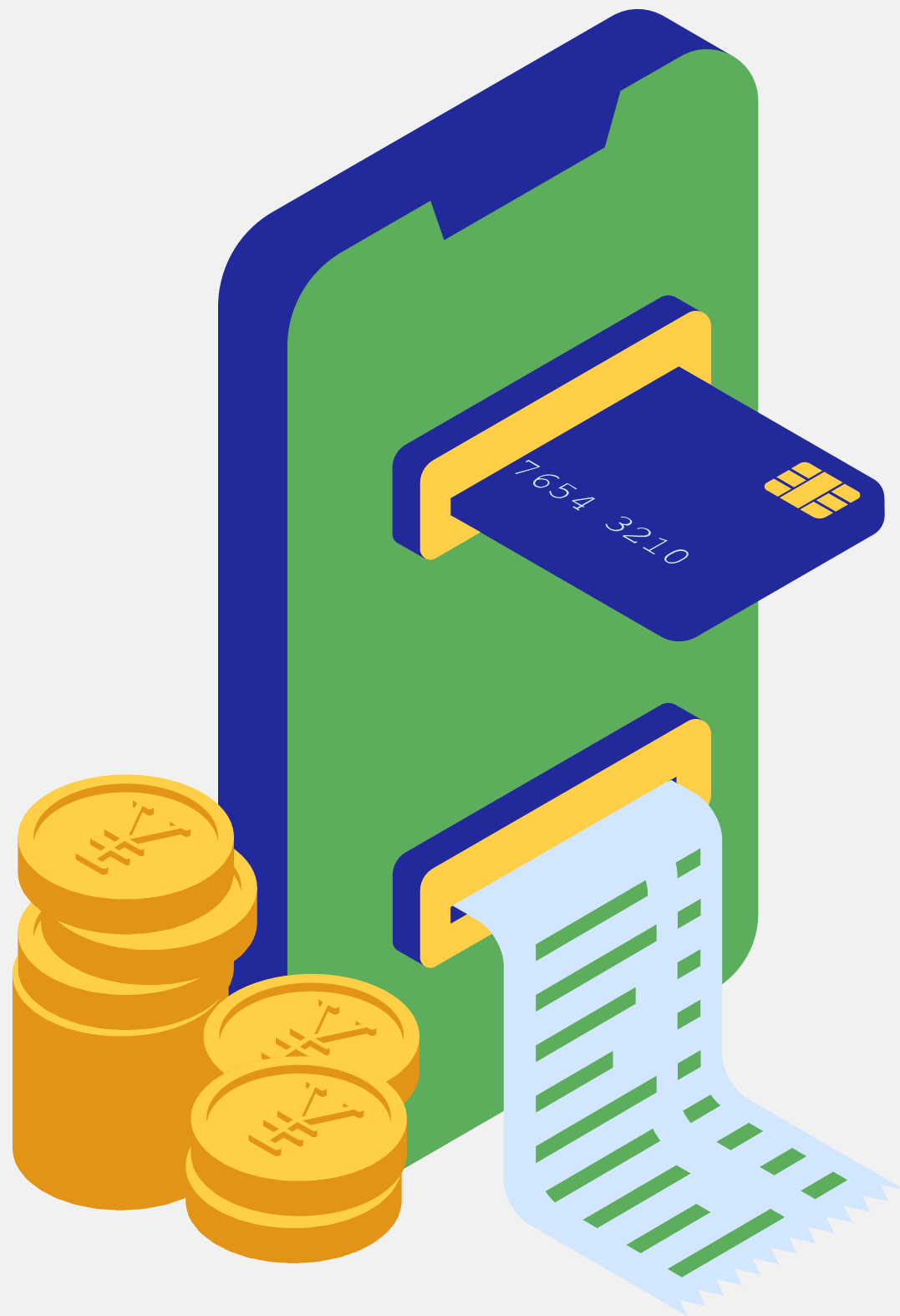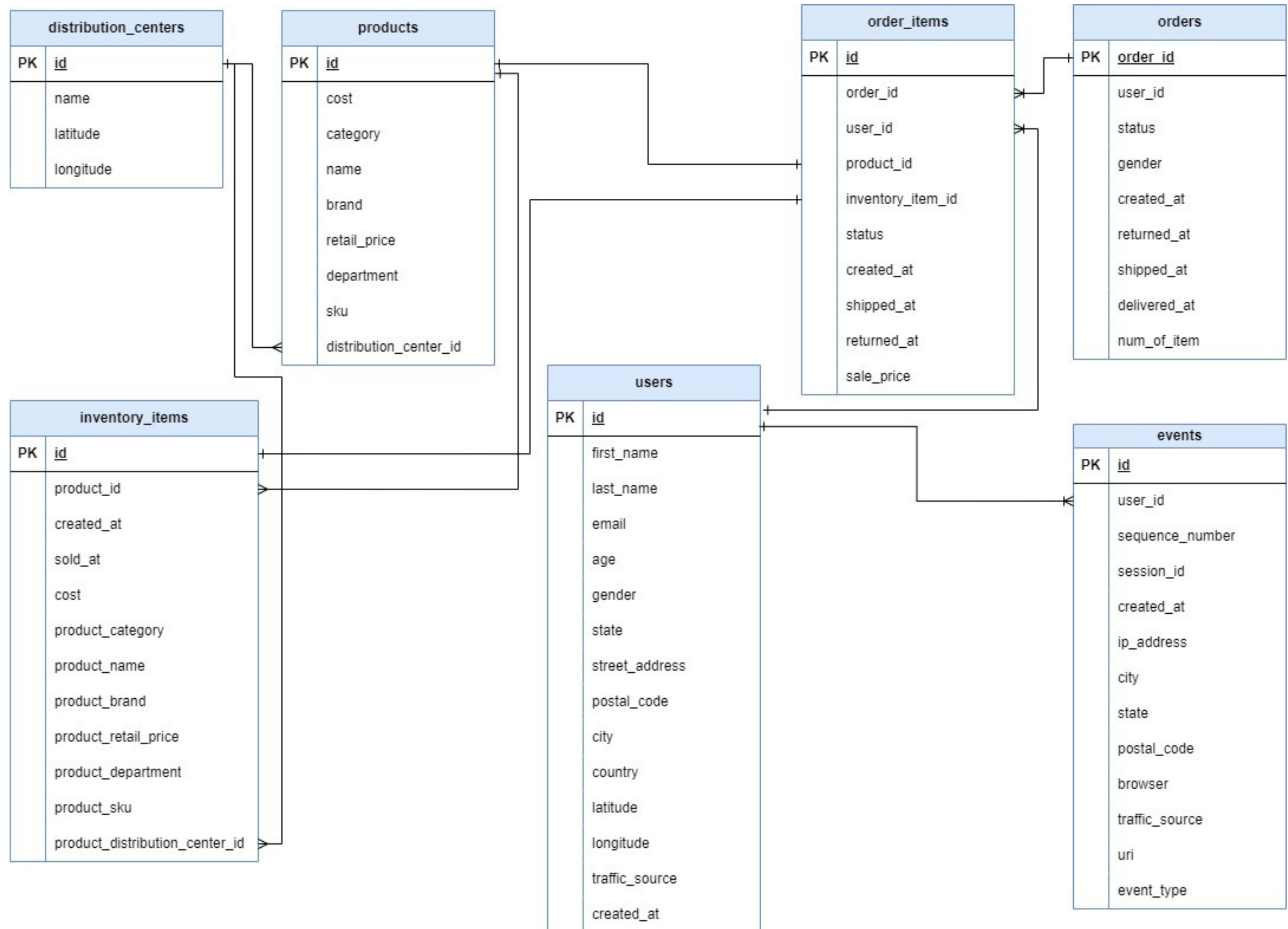
Introduction

TheLook is a fictitious eCommerce clothing site developed by the Looker team. The dataset contains information about customers, products, orders, logistics, and web events.

Database Stucture

distribution_centers

PK	id
	name
	latitude
	longitude

products

PK	id
	cost
	category
	name
	brand
	retail_price
	department
	sku
	distribution_center_id

order_items

PK	id
	order_id
	user_id
	product_id
	inventory_item_id
	status
	created_at
	shipped_at
	returned_at
	sale_price

orders

PK	order_id
	user_id
	status
	gender
	created_at
	returned_at
	shipped_at
	delivered_at
	num_of_item

inventory_items

PK	id
	product_id
	created_at
	sold_at
	cost
	product_category
	product_name
	product_brand
	product_retail_price
	product_department
	product_sku
	product_distribution_center_id

users

PK	id
	first_name
	last_name
	email
	age
	gender
	state
	street_address
	postal_code
	city
	country
	latitude
	longitude
	traffic_source
	created_at

events

PK	id
	user_id
	sequence_number
	session_id
	created_at
	ip_address
	city
	state
	postal_code
	browser
	traffic_source
	uri
	event_type

Objective

Insights from big data

The goal is to utilize SQL queries to delve into the dataset and extract valuable insights that can enhance business performance for theLook. This includes areas such as sales optimization, marketing effectiveness, and supply chain management efficiency.

Introduction

Customers rely more on e-commerce and the convenience it presents: shopping from home with just a few simple taps on one's device.

Its influence includes shaping a "buy-now" mindset and increased expectations of a clutter-free shopping experience.

CRM monthly report for 2024 Jan
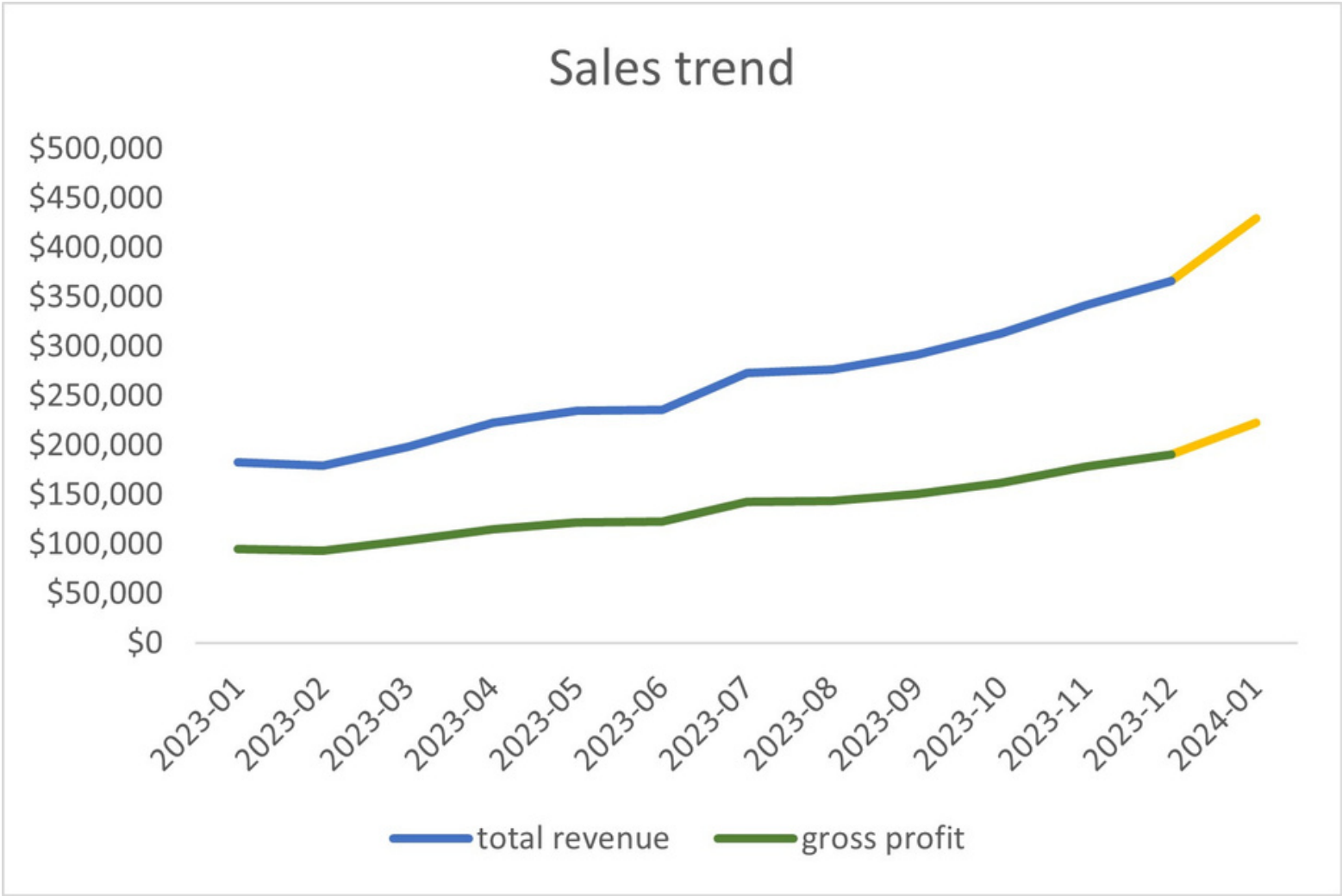
— Overall KPI

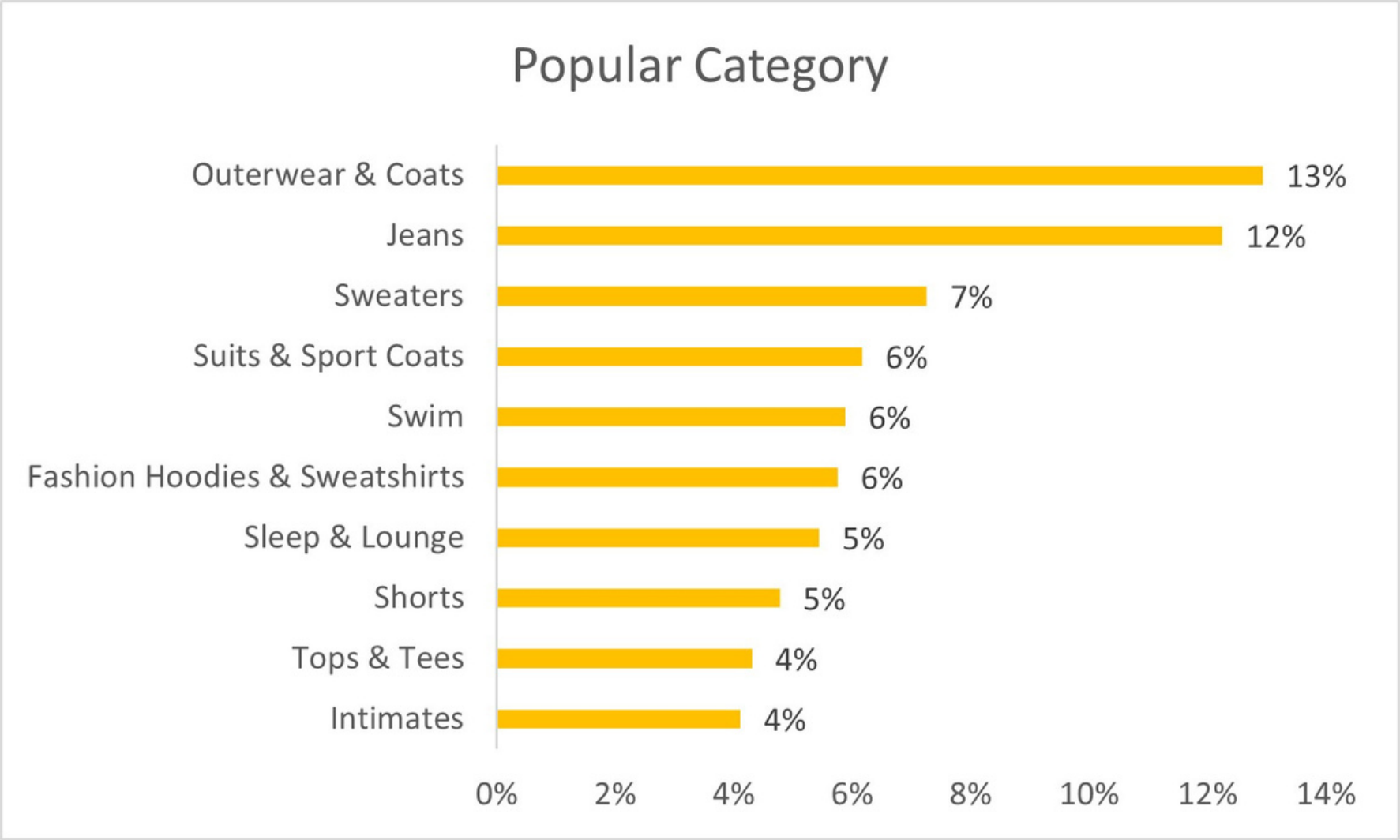**Total Revenue ($)**   **Total Cost ($)**   **Gross Profit ($)**

429,657.13   206,967.73   222,689.4

Sales trend

Revenue Compared with Last Year

Revenue Compared with Last Month
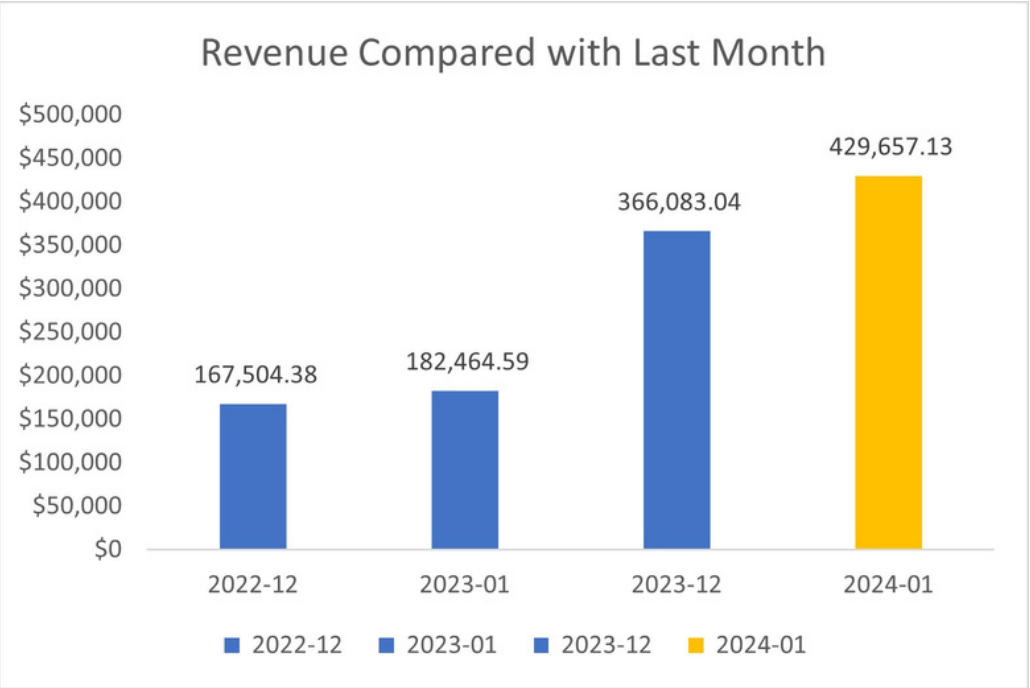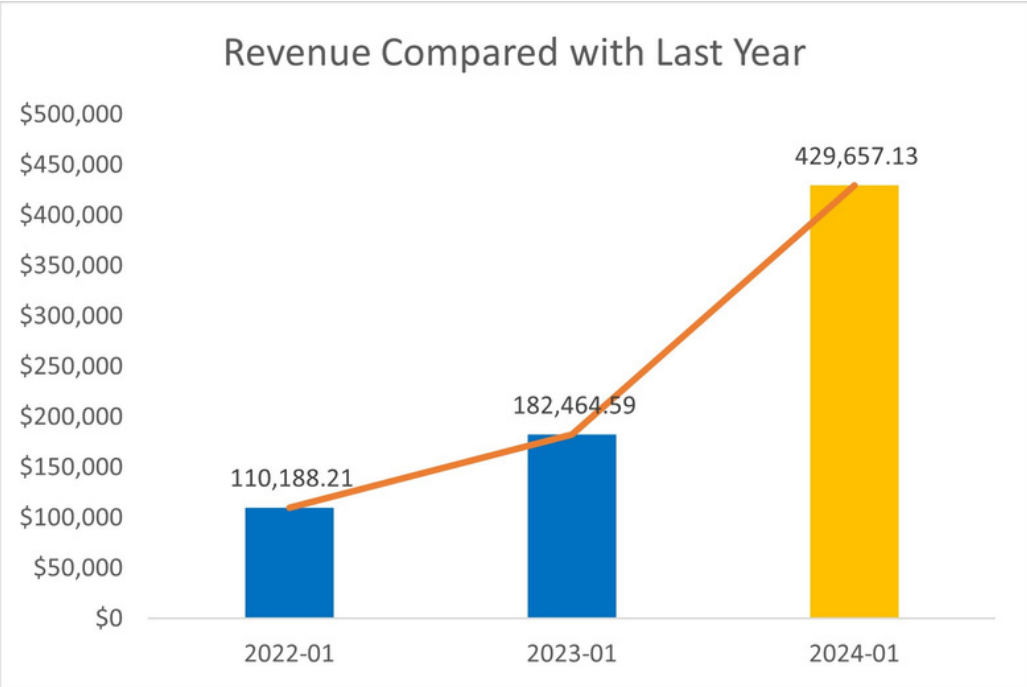
CRM monthly report for 2024 Jan

— Popular products

Popular Category

Category	Percentage
Outerwear & Coats	13%
Jeans	12%
Sweaters	7%
Suits & Sport Coats	6%
Swim	6%
Fashion Hoodies & Sweatshirts	6%
Sleep & Lounge	5%
Shorts	5%
Tops & Tees	4%
Intimates	4%

Name	Total Revenue	Total Cost	Gross Profit	Unit Sold
Diesel Men's Lisardo Jacket	1,516	670.07	845.93	2
Canada Goose Men's Lodge Jacket	1,200	532.8	667.2	3
Nicole Miller Women's Halter Jumpsuit	1,100	612.7	487.3	2
Alpha Industries Rip Stop Short	999	482.52	516.48	1
Nobis Yatesy Parka	950	381.9	568.1	1
The North Face Apex Bionic Soft Shell Jacket - Men's	903	391.9	511.1	1
Jordan Low Quarter Sock Style # 427411	903	537.29	365.71	1
The North Face Women's S-XL Oso Jacket	903	378.36	524.64	1
Mens Nike AirJordan Varsity Hoodie Jacket Grey / Black 451582-066	903	409.06	493.94	1
AIR JORDAN DOMINATE SHORTS MENS 465071-100	903	454.21	448.79	1

CRM monthly report for 2024 Jan

— Customer profile

Age Group Profile

Age Group	Value
Seniors	2526
Adults	1842
Young Adults	1927
Teens	847

Gender Profile

Gender	Value
Female	2453
Male	2390

Country Profile

Country	Value
China	1637
United States	1111
Brasil	738
South Korea	259
United Kingdom	227
France	208
Germany	195
Spain	179
Japan	106
Australia	102
Belgium	66
Poland	14
Colombia	1
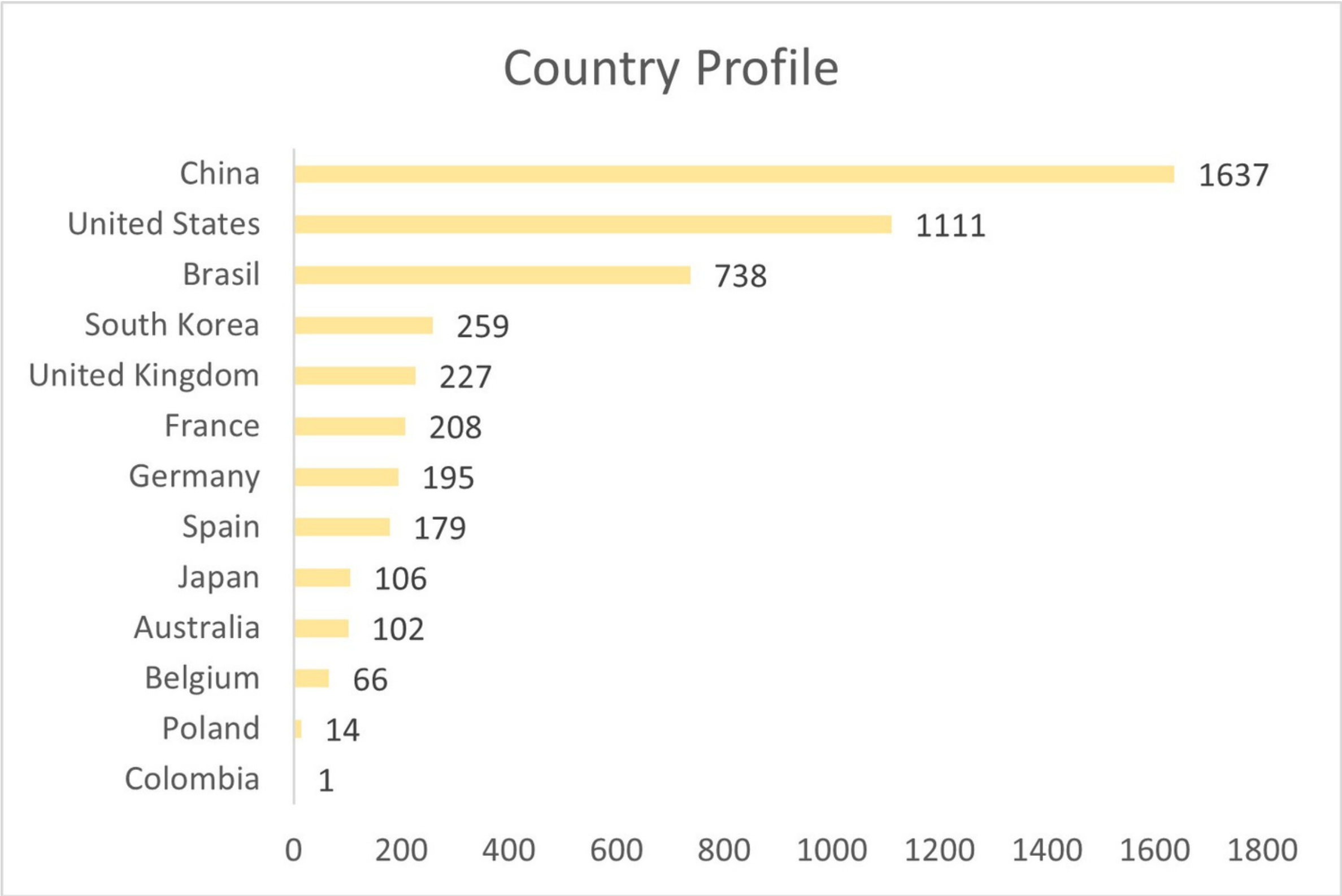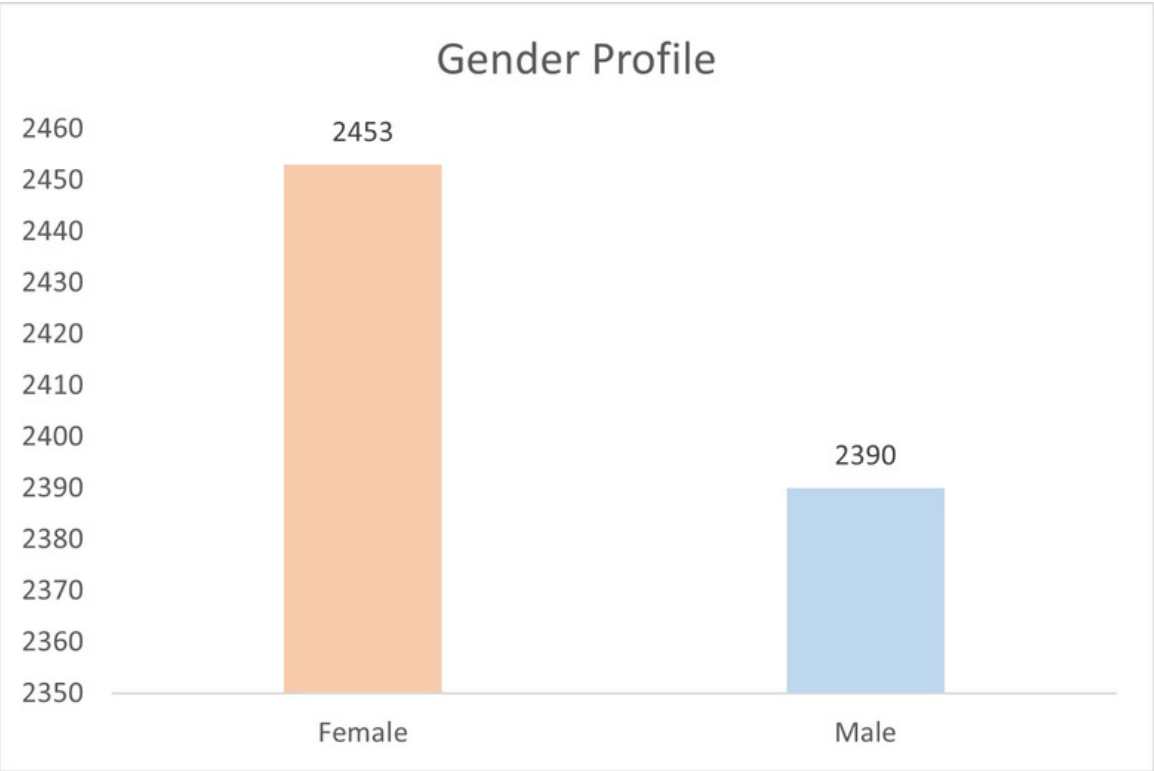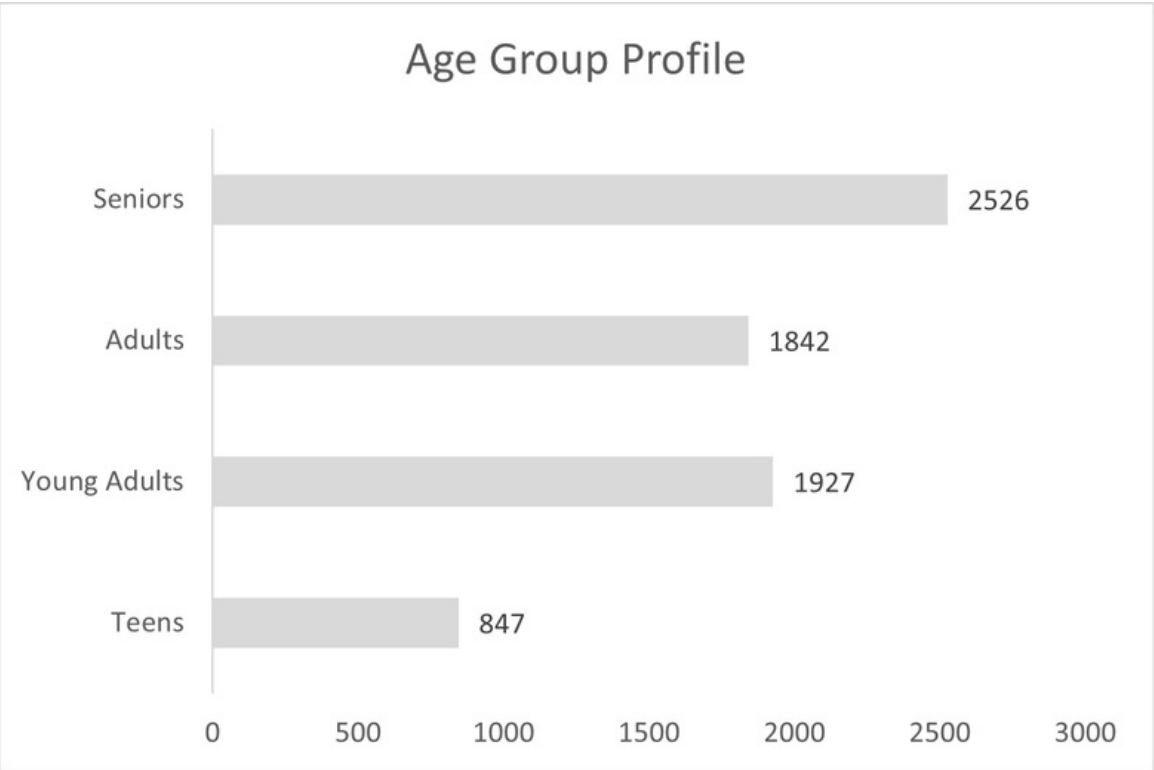
RFM Analysis 2019~2024

segment	recency	frequency	monetary	user_count	percent	user_id
Champions	27.49	2.65	167.26	2442	8.87	65423, 84596, 94492, 81887, 4104, 37155, 92227...
Loyal Customers	226.58	2.67	168.65	4511	16.39	11291, 97434, 49853, 64662, 21679, 64470, 9018...
Potential Loyalists	85.53	1.00	69.58	4158	15.11	6898, 11238, 15115, 16364, 17300, 29771, 33477...
New Customers	27.26	1.00	18.02	995	3.62	9618, 13757, 5462, 66952, 88431, 11937, 77055,...
Promising	139.06	1.00	17.95	1108	4.03	94582, 63005, 93890, 38293, 24855, 37516, 1871...
Need Attention	317.65	1.00	100.78	1073	3.90	43253, 44169, 57752, 62966, 64260, 71253, 7196...
About to Sleep	319.87	1.00	29.08	2234	8.12	87818, 929, 41414, 58741, 60485, 53355, 59797,...
At Risk	861.36	1.44	108.38	4486	16.30	38752, 40419, 42677, 43492, 47749, 48820, 5579...
Can't Lose	834.81	3.26	215.05	1895	6.89	62645, 65964, 62998, 14745, 4966, 15906, 6327,...
Hibernating	877.86	1.00	29.25	4617	16.78	28769, 19284, 11621, 20025, 70392, 20910, 2606...

Customers Segments

Potential Loyalists
85.5 days
1.0 orders
$69.6
4158 customers (15%)

Can't Lose
834.8 days
3.3 orders
$215.1
1895 customers (7%)

Hibernating
877.9 days
1.0 orders
$29.2
4617 customers (17%)

Loyal Customers
226.6 days
2.7 orders
$168.7
4511 customers (16%)

About to Sleep
319.9 days
1.0 orders
$29.1
2234 customers (8%)

At Risk
861.4 days
1.4 orders
$108.4
4486 customers (16%)

Champions
27.5 days
2.6 orders
$167.3
2442 customers (9%)

New Customers
27.3 days
1.0 orders
$18.0
995 customers (4%)

Promising
139.1 days
1.0 orders
$17.9
1108 customers (4%)

Need Attention
317.6 days
1.0 orders
$100.8
1073 customers (4%)

Distribution Center Analysis 2023

Manhattan distance is a metric in which the distance between two points is the sum of the absolute differences of their Cartesian coordinates. In a simple way of saying it is the total sum of the difference between the x-coordinates and y-coordinates.
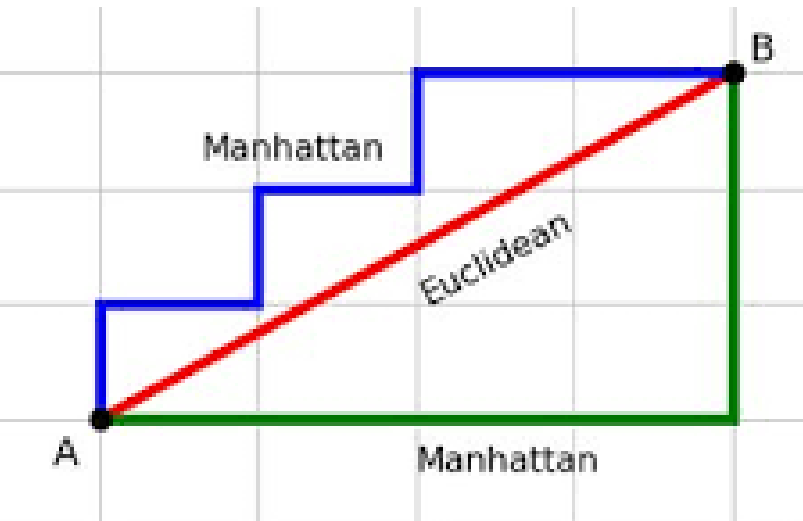
Description	Euclidean Distance (km)	Manhattan Distance (km)
Centroids distribution distance	1,194,912.93	1,688,942.69
Current DC distance	7,090,704.85	10,185,974.77
Distance difference km	5,895,791.92	8,497,032.08
% reduce	83.15 %	83.42 %

Customer Locations, Clusters, and Centroids in the US
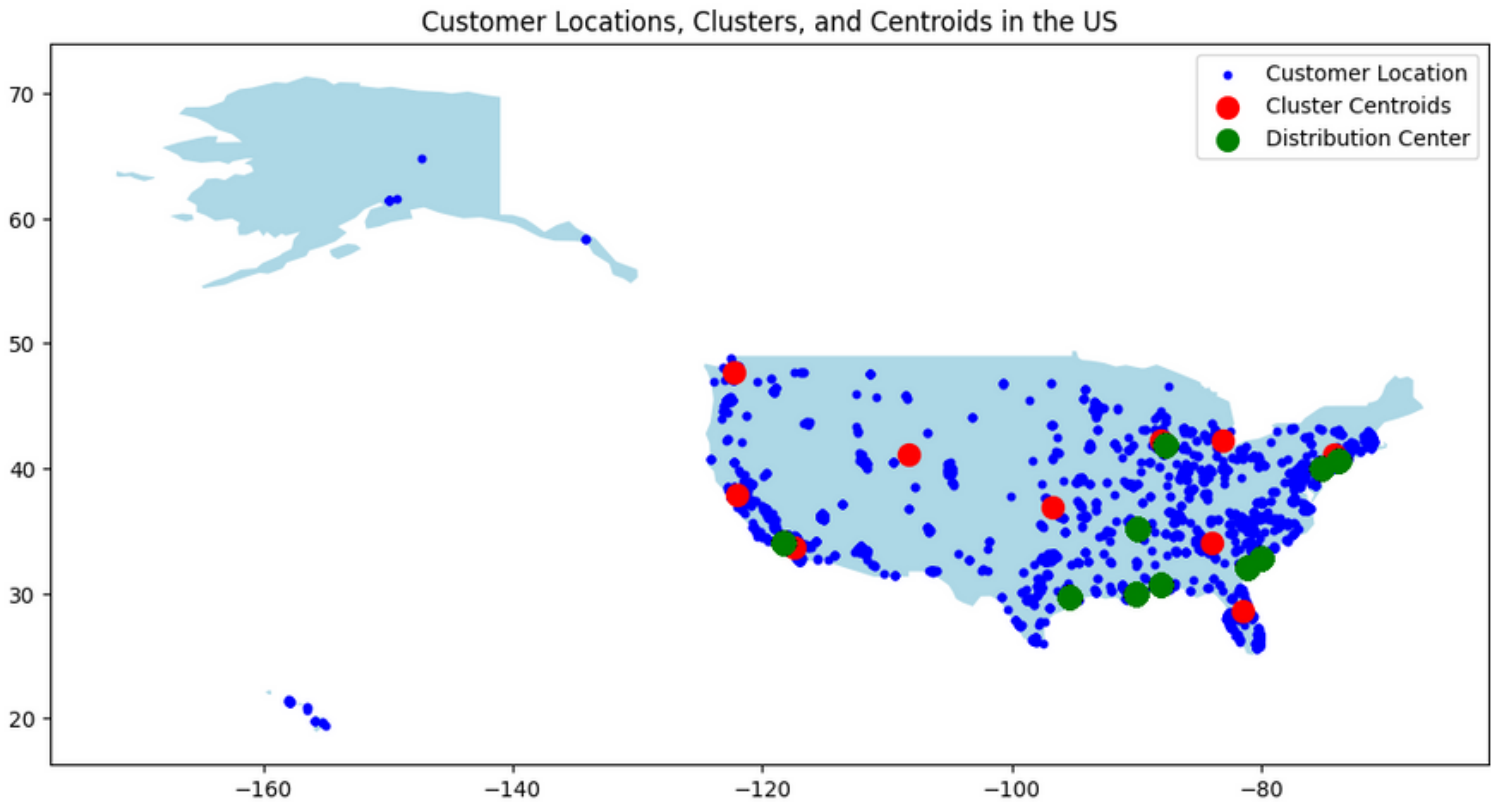
Time-series Analysis 2019/01~2024/03

The final forecast in Prophet is composed additively, combining different components:

$y(t)=g(t)+s(t)+h(t)+e(t)$

Where:

- $y(t)$ is the forecasted value at time t.
- $g(t)$ represents the trend component.
- $s(t)$ represents the seasonal component.
- $h(t)$ represents the holiday effect component.
- $e(t)$ represents the residual component, capturing uncertainty or unexplained variations.

This additive composition allows Prophet to model complex time series patterns by combining trends, seasonalities, holiday effects, and residual fluctuations.

Revenue Forecast with Prophet (Smooth Line, No Holiday Seasonality)
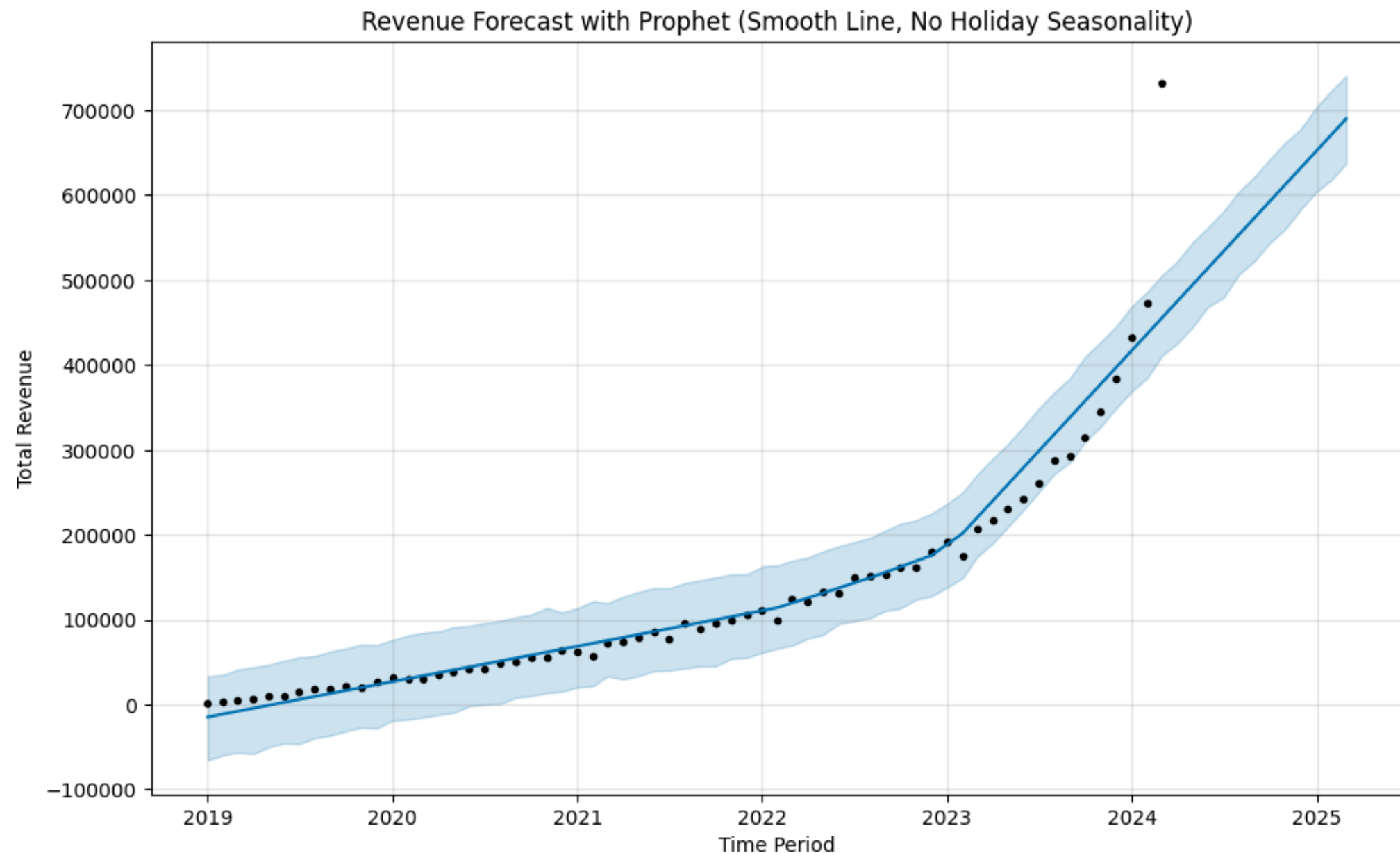
Limitations
& Future Prospectives

- Database design: some features are repetitive in different tables
- Database feature: more information should be incorporated
- Assumptions: synthetic data makes the data value dynamic
- Business value: lack of insider information to customize the business value
- Data model: for the time series model a more complex model can be developed to describe the pattern well

Recommendations

- Leveraged CRM monthly reports can identify the performance with ease

- RFM analysis benefits the customized marketing plan effectively and the matrics can be updated

- The location of distribution centers deserves to be reviewed and incorporated more business insider information to calculate the business value