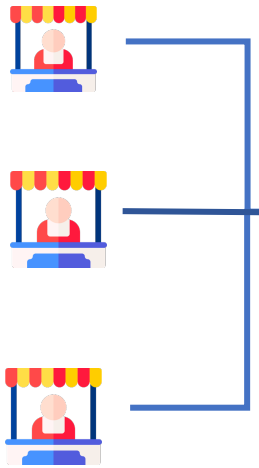
Building a Data Ecosystem in GCP for Scalable Trend Analysis

Created by: Rhythm Billore

Google Cloud Platform - Data Pipeline



Data Sources

Upload

CSV/Excel/
Parquet files

Web Portal Development
(Python Flask)

Google Cloud Storage (GCS)
(store files in bucket)

Trigger message
(new file available)

Trigger message
(new file uploaded)

Google Pub/Sub

Pub/Sub Acknowledgment

Sends trigger

Google Cloud Function

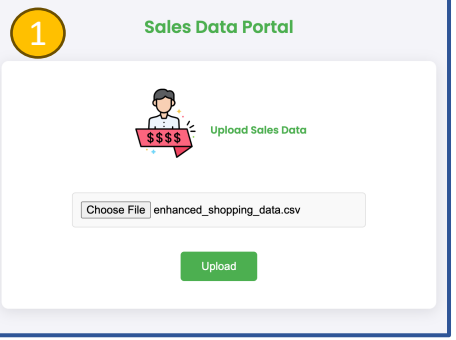
Load data

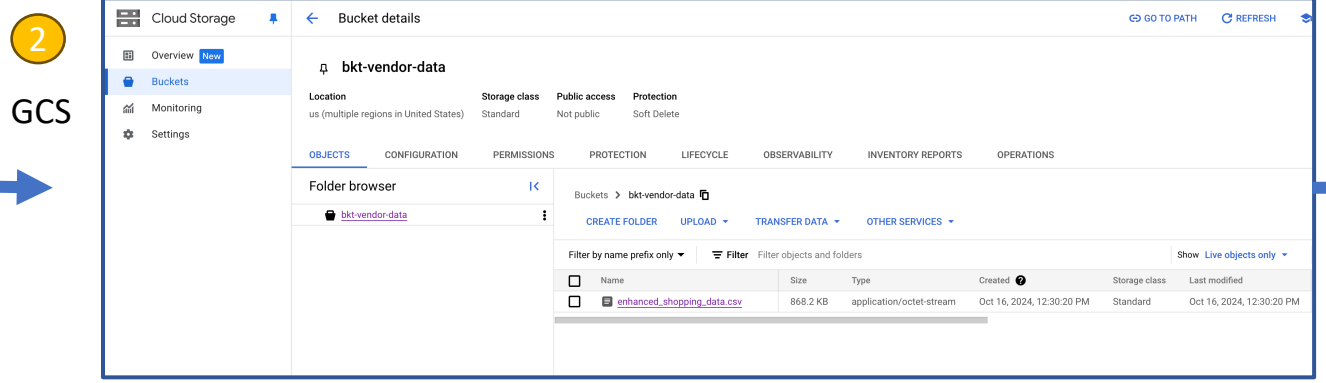
Big Query

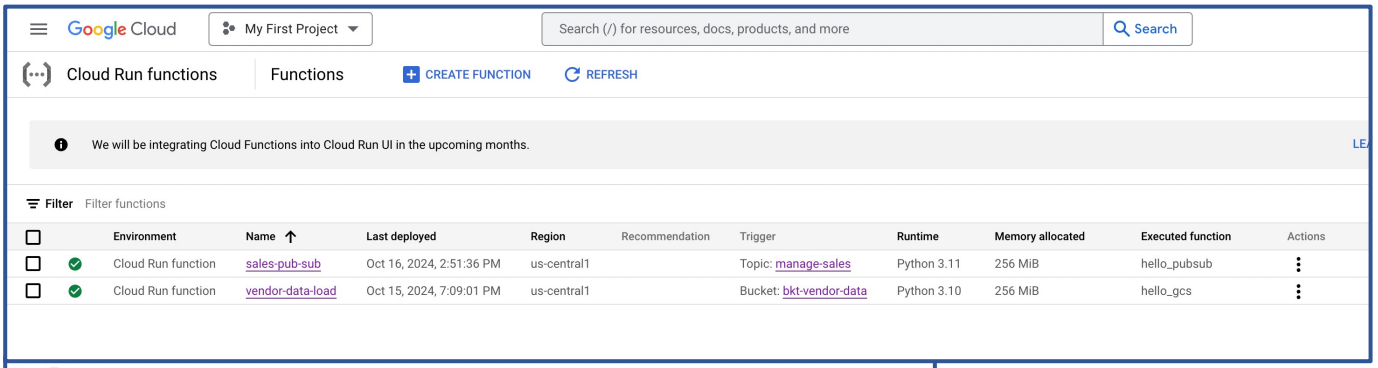
Looker

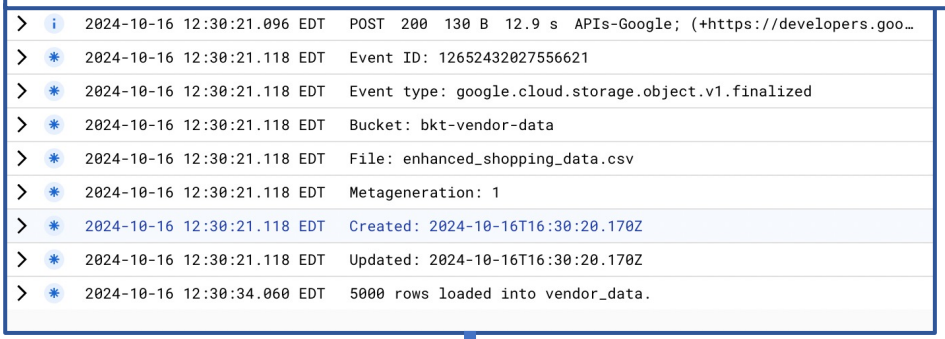
vertex.ai

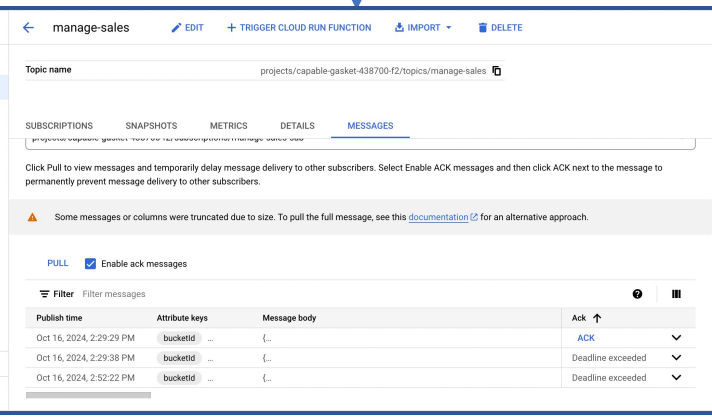
Business Users


1
Sales Data Portal
Upload Sales Data
$$$$
Choose File
enhanced_shopping_data.csv
Upload

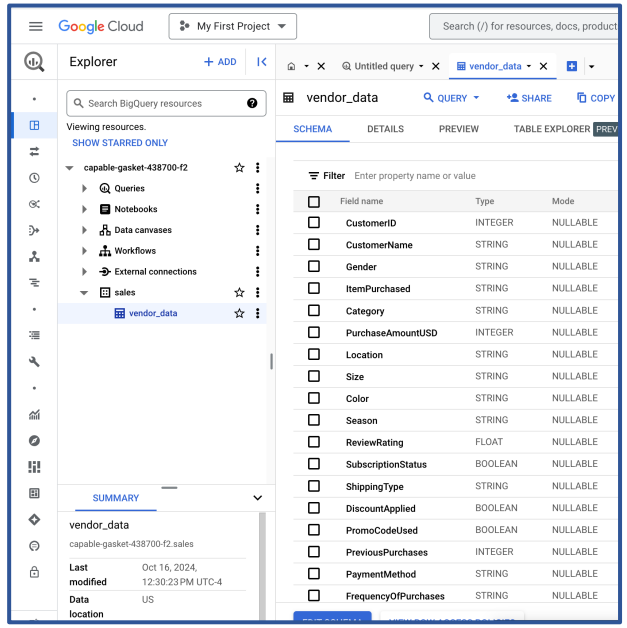

2
GCS
Cloud Storage
Overview
New
Buckets
Monitoring
Settings
Bucket details
GO TO PATH
REFRESH
bkt-vendor-data
Location
us (multiple regions in United States)
Storage class
Standard
Public access
Not public
Protection
Soft Delete
OBJECTS
CONFIGURATION
PERMISSIONS
PROTECTION
LIFECYCLE
OBSERVABILITY
INVENTORY REPORTS
OPERATIONS
Folder browser
bkt-vendor-data
Buckets
bkt-vendor-data
CREATE FOLDER
UPLOAD
TRANSFER DATA
OTHER SERVICES
Filter by name prefix only
Filter
Filter objects and folders
Show
Live objects only
Name
Size
Type
Created
Storage class
Last modified
enhanced_shopping_data.csv
868.2 KB
application/octet-stream
Oct 16, 2024, 12:30:20 PM
Standard
Oct 16, 2024, 12:30:20 PM


Google Cloud
My First Project
Search (/) for resources, docs, products, and more
Search
Cloud Run functions
Functions
CREATE FUNCTION
REFRESH
We will be integrating Cloud Functions into Cloud Run UI in the upcoming months.
Filter
Filter functions
Environment
Name
Last deployed
Region
Recommendation
Trigger
Runtime
Memory allocated
Executed function
Actions
Cloud Run function
sales-pub-sub
Oct 16, 2024, 2:51:36 PM
us-central1
Topic: manage-sales
Python 3.11
256 MiB
hello_pubsub
Cloud Run function
vendor-data-load
Oct 15, 2024, 7:09:01 PM
us-central1
Bucket: bkt-vendor-data
Python 3.10
256 MiB
hello_gcs


2024-10-16 12:30:21.096 EDT POST 200 130 B 12.9 s APIs-Google; (+https://developers.goo…
2024-10-16 12:30:21.118 EDT Event ID: 12652432027556621
2024-10-16 12:30:21.118 EDT Event type: google.cloud.storage.object.v1.finalized
2024-10-16 12:30:21.118 EDT Bucket: bkt-vendor-data
2024-10-16 12:30:21.118 EDT File: enhanced_shopping_data.csv
2024-10-16 12:30:21.118 EDT Metageneration: 1
2024-10-16 12:30:21.118 EDT Created: 2024-10-16T16:30:20.170Z
2024-10-16 12:30:21.118 EDT Updated: 2024-10-16T16:30:20.170Z
2024-10-16 12:30:34.060 EDT 5000 rows loaded into vendor_data.


3
Pub/Sub
Pub/Sub
Topics
Subscriptions
Snapshots
Schemas
Pub/Sub Lite
Lite Reservations
Lite Topics
Lite Subscriptions
Release Notes
manage-sales
EDIT
TRIGGER CLOUD RUN FUNCTION
IMPORT
DELETE
Topic name
projects/capable-gasket-438700-f2/topics/manage-sales
SUBSCRIPTIONS
SNAPSHOTS
METRICS
DETAILS
MESSAGES
Click Pull to view messages and temporarily delay message delivery to other subscribers. Select Enable ACK messages and then click ACK next to the message to permanently prevent message delivery to other subscribers.
Some messages or columns were truncated due to size. To pull the full message, see this documentation for an alternative approach.
PULL
Enable ack messages
Filter
Filter messages
Publish time
Attribute keys
Message body
Ack
Oct 16, 2024, 2:29:29 PM
bucketId
ACK
Oct 16, 2024, 2:29:38 PM
bucketId
Deadline exceeded
Oct 16, 2024, 2:52:22 PM
bucketId
Deadline exceeded

4
Cloud
function


Google Cloud
My First Project
Search (/) for resources, docs, product
Explorer
ADD
Search BigQuery resources
Viewing resources.
SHOW STARRED ONLY
capable-gasket-438700-f2
Queries
Notebooks
Data canvases
Workflows
External connections
sales
vendor_data
Untitled query
vendor_data
vendor_data
QUERY
SHARE
COPY
SCHEMA
DETAILS
PREVIEW
TABLE EXPLORER
Filter
Enter property name or value
Field name
Type
Mode
CustomerID
INTEGER
NULLABLE
CustomerName
STRING
NULLABLE
Gender
STRING
NULLABLE
ItemPurchased
STRING
NULLABLE
Category
STRING
NULLABLE
PurchaseAmountUSD
INTEGER
NULLABLE
Location
STRING
NULLABLE
Size
STRING
NULLABLE
Color
STRING
NULLABLE
Season
STRING
NULLABLE
ReviewRating
FLOAT
NULLABLE
SubscriptionStatus
BOOLEAN
NULLABLE
ShippingType
STRING
NULLABLE
DiscountApplied
BOOLEAN
NULLABLE
PromoCodeUsed
BOOLEAN
NULLABLE
PreviousPurchases
INTEGER
NULLABLE
PaymentMethod
STRING
NULLABLE
FrequencyOfPurchases
STRING
NULLABLE
SUMMARY
vendor_data
capable-gasket-438700-f2.sales
Last modified
Oct 16, 2024, 12:30:23 PM UTC-4
Data location
US

5
Big Query

Sales Data Analysis - 2024

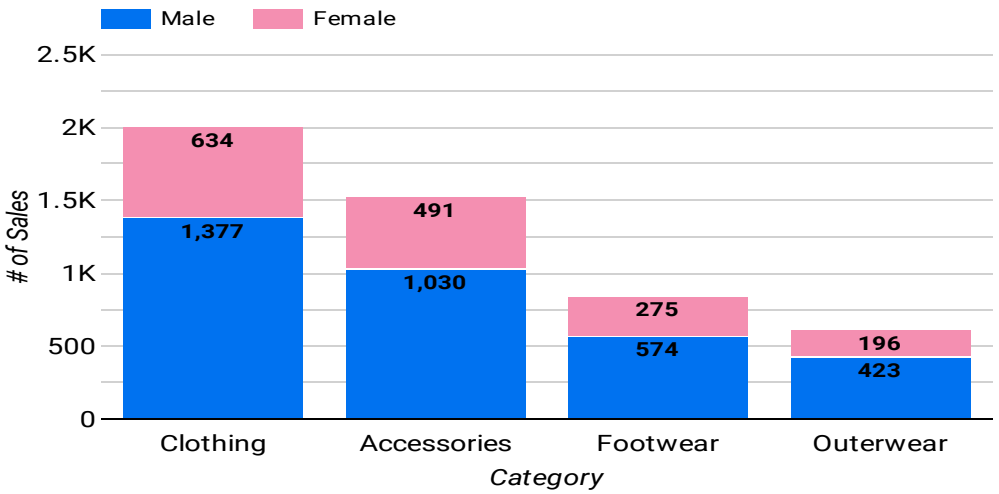
# of Sales
5000

States Covered
50

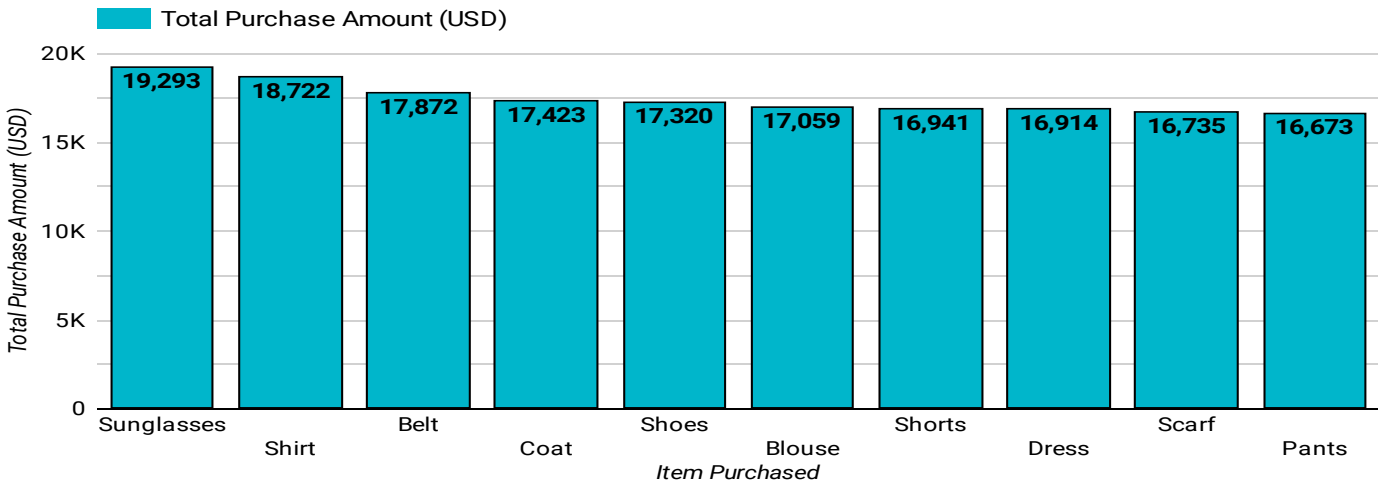
Net Revenue
402.2K

# of Active Subscribers
1,343

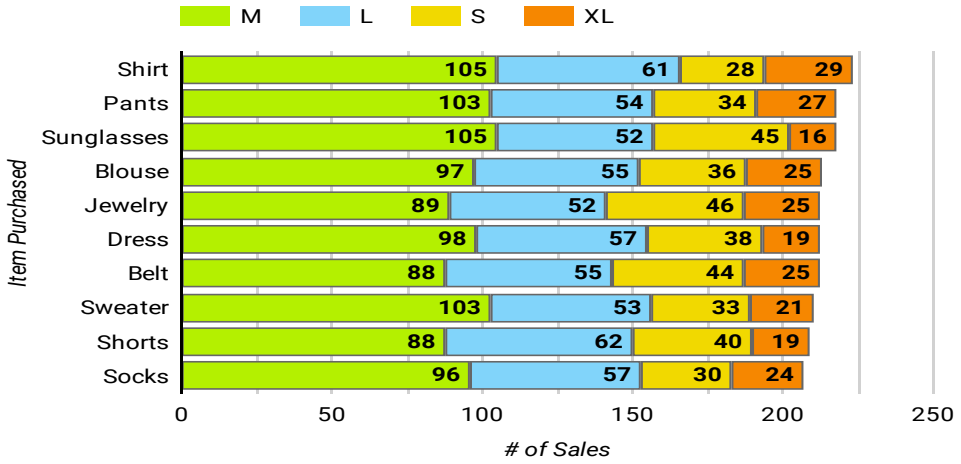
Sales Across Categories



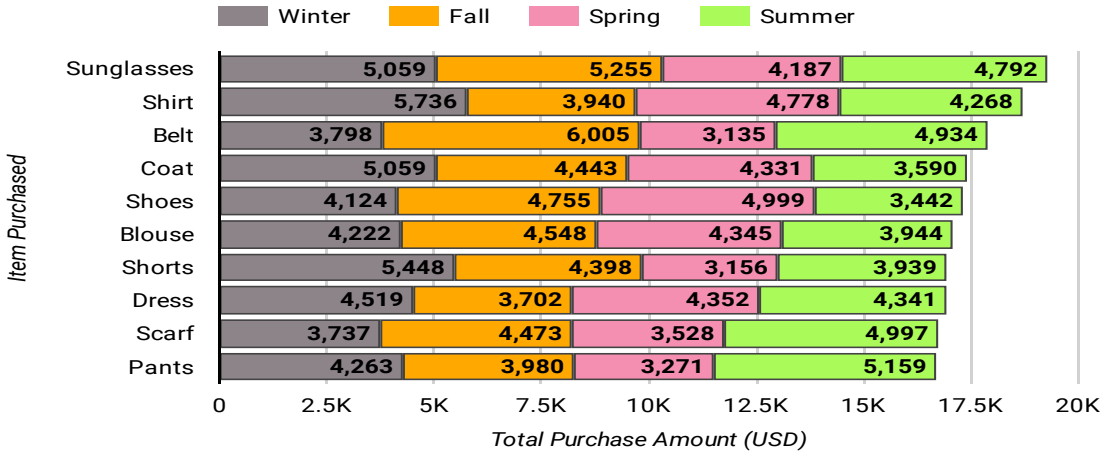
Top 10 Selling Items



Sales Trend by Size

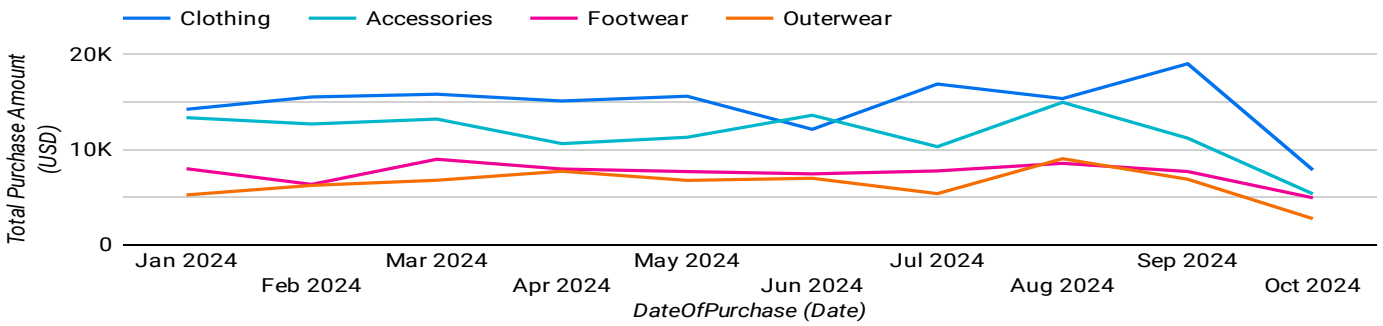


Season-wise Sales Trend

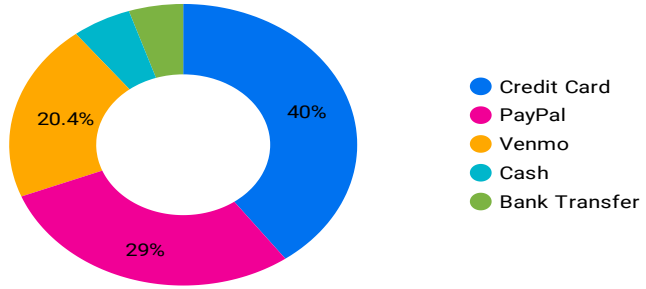


Customer Behavior analysis - 2024

Month-on-Month Purchase Trends by Category (2024)

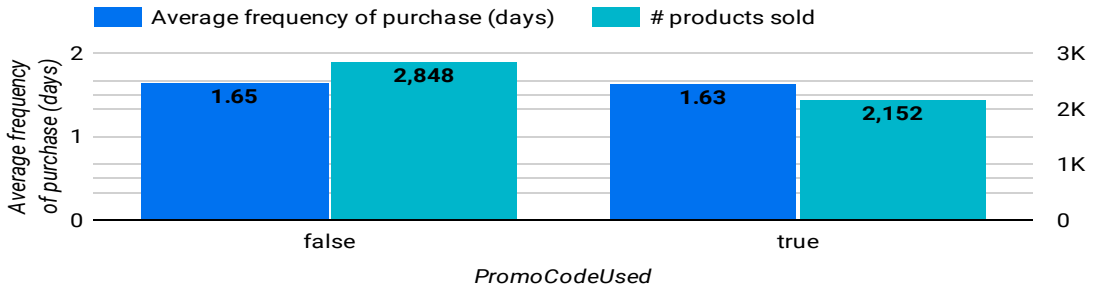


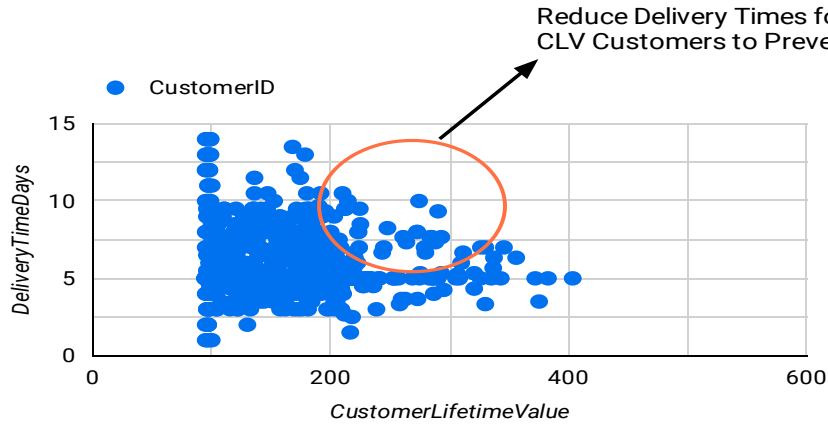
Transaction Share by Payment Method



Promo Code Usage vs. Purchase Frequency

(Promo codes applied on 85 / 295 days)





	Loyalty Segment	SubscriptionStatus	Current Status	Conversion Target	No. of Customers	Frequency of purchase (in days)	Avg Purchase Amount (USD)	CustomerLifetimeValue
1.	Low Loyalty Points	false	One-Off Buyers	High Loyality	3,135	1.62	59.9	68.2
2.	Low Loyalty Points	true			1,141	1.68	59.1	65.07
3.	Medium Loyalty Points	false			273	1.63	160.9	236.83
4.	High Loyalty Points	false	Loyal Non-Subscribers	Subscribers	247	1.7	256.2	333.24
5.	Medium Loyalty Points	true			105	1.76	154.8	241.54
6.	High Loyalty Points	true	Subscribers	Retain	95	1.39	246.9	335.02

1 - 8 / 8