# GOVERNMENT COLLEGE OF ENGINEERING ERODE

அரசினர் பொறியியல் கல்லூரி, ஈரோடு
Government College of Engineering, Erode
(Approved by AICTE, New Delhi and Affiliated to Anna University, Chennai)

B.E Electronics and Communication Engineering

## PRODUCT SALES ANALYSIS


**Name of the Student:** KATHIRIPERUMAL .S

**University Register no:** 731121106024

Under the mentor of

**Dr.M.Poongothai**

**Department of Information Technology(IT)**

**Department of Electronics and Communication Engineering**

Government College of Engineering

Erode ,PO ,near Vasavi College,TamilNadu-638316,

Affiliated to Anna University ,Chennai.

# PRODUCT SALES ANALYSIS

## AIM:

To find the **Region** wise top 5 and bottom 5 products with **Order_ID** by considering the **Profit** as the main aspect.

## PROJECT DEFINITION:

Product sales analysis is the process of examining and evaluating data related to the sales of a particular product or a group of products within a business. The primary goal of product sales analysis is to gain insights into how well a product is performing in the market and to make informed decisions to improve sales and profitability.

## PROJECT OVERVIEW:

The Product Sales Analysis Data Analytics Project with Cognos is a strategic initiative aimed at leveraging the power of IBM Cognos, a robust business intelligence and analytics platform, to gain deep insights into a company's product sales performance. This project involves the systematic collection, processing, analysis, and visualization of sales data to drive data-driven decision-making, enhance business profitability, and optimize product-related strategies.

## PROJECT OBJECTIVES:

### Sales Performance Assessment:

The primary objective of this project is to assess the performance of products in terms of Region , Product name and Profit. This involves a comprehensive analysis of historical sales data.

### Customer Segmentation:

Utilize Cognos to segment customers based on demographics, purchasing behavior, and geography. Understanding customer segments helps tailor marketing strategies and product offerings.

### Trend Analysis:

Identify sales trends, seasonality, and cyclical patterns using historical data. This enables the business to anticipate and plan for future sales fluctuations.

### Competitive Benchmarking:

Compare the company's product sales with those of competitors. Identify strengths, weaknesses, opportunities, and threats in the market.

### Inventory Management:

Optimize inventory levels by analyzing sales trends and forecasting future demand. Prevent overstocking or understocking of products.

### Pricing Strategy Optimization:

Use data analytics to determine the most effective pricing strategies that maximize revenue while maintaining competitiveness.

**Marketing Campaign Effectiveness:**

Analyze the impact of marketing campaigns on product sales. Identify which campaigns generated the highest ROI and customer engagement.

## ABOUT DATASET :

- Order_ID : A specific ID given to each product
- Order_Priority : Priority of the product
- Order_Quantity : No of product items sold.
- Ship_Mode : Divided in two categories - Express Air and Regular Air
- Profit : Profit earned from the sale
- Customer_Name : Name of the customer purchasing the products
- Region : Region to which the customer belongs
- Customer_Segment : Divided as per the size of business
- Product_Category : Divided according to the usage of the product
- Product_Sub-Category : Divided according to the usage of the product
- Product_Name : Name of the product
- Product_Container : Type of container in which the product is shipped.

# MACHINE LEARNING

## DEFINITION OF MACHINE LEARNING :

Machine learning is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed. Machine learning systems use data and statistical techniques to improve their performance on a specific task over time. The primary goal of machine learning is to develop algorithms that can generalize from data, allowing them to make accurate predictions or decisions on new, unseen data.

## LINEAR REGRESSION : [Supervised]

Linear regression is one of the fundamental techniques in machine learning and statistics used for modeling the relationship between a dependent variable (target) and one or more independent variables (features or predictors). It is a type of supervised learning algorithm that aims to find the linear relationship between the features and the target variable. Linear regression is widely used for tasks such as predicting house prices, estimating sales figures, and understanding the relationship between variables in various fields.

## IMPLEMENTATION AND WORKING OF THE LINEAR REGRESSION :

- **Gather Data :**
  Collect and organize the data you'll use for training and testing, ensuring it's relevant to the problem you're addressing.
- **Preprocess Data :**
  Clean, handle missing values, and preprocess the data to make it suitable for training.
- **Split Data :**
  Divide the dataset into training and testing sets to evaluate the model's performance.
- **Choose Model :**
  Select linear regression as the model for the problem, a simple yet powerful choice for regression tasks.
- **Train Model :**
  Use the training data to fit the linear regression model, adjusting its parameters to minimize the error.
- **Evaluate Model :**
  Assess the model's performance using the testing data, typically using metrics like mean squared error (MSE) or R-squared.
- **Tune and Optimize :**
  Adjust hyperparameters or explore feature engineering to improve the model's performance.
- **Predict :**
  Use the trained model to make predictions on new data, providing valuable insights or predictions based on the trained linear regression.

## PREPROCESSING IN DATA ANALYTICS :

Preprocessing is a crucial step in data analytics and data science. It involves a series of operations that are performed on raw data to clean, transform, and prepare it for analysis. The quality of the preprocessing phase significantly impacts the accuracy and effectiveness of any data analytics or machine learning project. Here are some common preprocessing steps in data analytics:

- **Data Collection:**
  This is the initial step where data is gathered from various sources, such as databases, files, APIs, or sensors.
- **Handling Missing Values:**
  Identify and deal with missing data, which can involve imputation (filling in missing values) or removal of incomplete records.
- **Outlier Detection:**
  Identify and handle outliers, which are data points significantly different from the rest and can skew results.
- **Data Encoding:**
  Convert categorical variables into numerical representations (e.g., one-hot encoding or label encoding).
- **Feature Scaling:**

Standardize or normalize numerical features to have a common scale (e.g., using z-score normalization or min-max scaling).

- **Data Integration:**
  Combine data from multiple sources or datasets if necessary.
- **Data Splitting:**
  Divide the dataset into training, validation, and test sets for model training, evaluation, and testing.
- **Data Visualization:**
  Generate visualizations to explore data and gain insights into its characteristics, relationships, and potential issues.
- **Data Normalization:**
  Ensure that data follows a normal distribution, if required, to meet assumptions of certain statistical tests or machine learning algorithms.
- **Data Scaling:**
  Scale features to have similar ranges to avoid bias in some machine learning algorithms.
- **Text Preprocessing:**
  For natural language processing (NLP) tasks, this might include tokenization, stemming, lemmatization, and removing stop words.

## PYTHON IMPLEMENTATION:

### Importing relevant libraries

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Scikit-learn

## LINEAR REGRESSION PYTHON CODE :

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer

# Load and preprocess the sales data
data = pd.read_csv("D:\IBM PROJECT\productsalesanalysis1.csv")

# Select the features and target variable
X = data[['Product_Name', 'Region', 'Customer_Segment', 'Product_Category', 'Product_Sub-Category', 'Sales', 'Product_Container', 'Ship_Mode']]
Y = data['Profit']
```

```
# Specify which features are categorical
categorical_features = ['Product_Name', 'Region', 'Customer_Segment', 'Product_Category',
'Product_Sub-Category', 'Product_Container', 'Ship_Mode']

# One-hot encode categorical features
ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(), categorical_features)],
remainder='passthrough')
X = ct.fit_transform(X)

# Avoiding the dummy variable trap by dropping one column for each one-hot encoded feature
X = X [ : , 1 :]

# Split the data into training and testing sets
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42)

# Train the linear regression model
model = LinearRegression()
model.fit(X_train, Y_train)

# Ask the user for input: Region
region_input = input("Enter a Region: ")

# Filter the dataset for the specified region
filtered_data = data[data['Region'] == region_input]

# If there are no records for the specified region, inform the user
if filtered_data.shape[0] == 0:
    print("No records found for the specified region.")
else:
# Use the model to predict profit for the filtered data
X_filtered = filtered_data[['Product_Name', 'Region', 'Customer_Segment', 'Product_Category',
'Product_Sub-Category', 'Sales', 'Product_Container', 'Ship_Mode']]
 X_filtered = ct.transform(X_filtered)
 X_filtered = X_filtered[:, 1:]
 y_pred_filtered = model.predict(X_filtered)

 # Sort the filtered data by predicted profit in descending order
 sorted_data = filtered_data.copy()
 sorted_data['Predicted_Profit'] = y_pred_filtered
    sorted_data = sorted_data.sort_values(by='Predicted_Profit', ascending=False)

    # Get the top 5 and bottom 5 records
    top_5 = sorted_data.head(5)
    bottom_5 = sorted_data.tail(5)

    # Print the top 5 predicted profit values and corresponding Order_IDs, Product Names
    print("\nTop 5 Predicted Profit Values:")
```

```
for index, row in top_5.iterrows():
print(f"Order_ID: {row['Order_ID']}, Product Name: {row['Product_Name']},
Predicted Profit:{row['Predicted_Profit']:.2f}")

# Print the bottom 5 predicted profit values and corresponding Order_IDs, Product Names
print("\nBottom 5 Predicted Profit Values:")
for index, row in bottom_5.iterrows():
print(f"Order_ID: {row['Order_ID']}, Product Name: {row['Product_Name']},
Predicted Profit: {row['Predicted_Profit']:.2f}")

# Create a scatter plot for visualization
plt.figure(figsize=(8, 6))
plt.scatter(filtered_data['Profit'], y_pred_filtered, color='blue', marker='o')
plt.title('Actual vs. Predicted Profit for the Specified Region')
plt.xlabel('Actual Profit')
plt.ylabel('Predicted Profit')

# Add a diagonal line for reference (perfect predictions)
plt.plot([filtered_data['Profit'].min(), filtered_data['Profit'].max()],
[filtered_data['Profit'].min(), filtered_data['Profit'].max()], 'k--', lw=2)

# Show the plot
plt.show()
```

**OUTPUT :**

**Region Wise Top 5 And Bottom 5 Product Names with Order_ID by Profit :**

**Output 1:**

Enter a Region: **Ontario**

**Top 5** Predicted Profit Values:

Order_ID: 41059, Product Name: Polycom VoiceStation 100, Predicted Profit: 5457.90

Order_ID: 33250, Product Name: Kensington 7 Outlet MasterPiece Power Center, Predicted Profit: 2494.60

Order_ID: 31238, Product Name: Hoover WindTunnel Plus Canister Vacuum, Predicted Profit: 2259.21

Order_ID: 3109, Product Name: Talkabout T8097, Predicted Profit: 1505.16

Order_ID: 56101, Product Name: GBC Therma-A-Bind 250T Electric Binding System, Predicted Profit: 1344.25

**Bottom 5** Predicted Profit Values:

Order_ID: 51780, Product Name: Fellowes Twister Kit, Gray/Clear, 3/pkg, Predicted Profit: -226.33

Order_ID: 33255, Product Name: Xerox 4200 Series MultiUse Premium Copy Paper (20Lb. and 84 Bright), Predicted Profit: -301.81

Order_ID: 28290, Product Name: Fellowes Super Stor/Drawer  Files, Predicted Profit: -428.79

Order_ID: 29318, Product Name: SAFCO Commercial Wire Shelving, Black, Predicted Profit: -640.77

Order_ID: 50533, Product Name: Space Solutions Commercial Steel Shelving, Predicted Profit: -900.20

**Output 2:**

Enter a Region: **Nunavut**

**Top 5** Predicted Profit Values:

Order_ID: 7110, Product Name: SAFCO Arco Folding Chair, Predicted Profit: 1724.96

Order_ID: 483, Product Name: R380, Predicted Profit: 1253.74

Order_ID: 7430, Product Name: #10 White Business Envelopes,4 1/8 x 9 1/2, Predicted Profit: 389.30

Order_ID: 1344, Product Name: LX 788, Predicted Profit: 380.22

Order_ID: 7906, Product Name: 3M Office Air Cleaner, Predicted Profit: 297.28

**Bottom 5** Predicted Profit Values:

Order_ID: 6916, Product Name: Crate-A-Files , Predicted Profit: -90.11

Order_ID: 1539, Product Name: GBC Pre-Punched Binding Paper, Plastic, White, 8-1/2" x 11", Predicted Profit: -124.11

Order_ID: 643, Product Name: SAFCO Commercial Wire Shelving, Black, Predicted Profit: -606.72

Order_ID: 4612, Product Name: Hoover Portapower Portable Vacuum, Predicted Profit: -1157.06

Order_ID: 6116, Product Name: High Speed Automatic Electric Letter Opener, Predicted Profit: -1876.84

**Output 3:**

Enter a Region: **Northwest Territories**

**Top 5** Predicted Profit Values:

Order_ID: 21383, Product Name: Polycom ViaVideo Desktop Video Communications Unit, Predicted Profit: 7432.31

Order_ID:12419,ProductName:PolycomViewStationAdapterH323 Videoconferencing Unit, Predicted Profit: 5309.54

Order_ID: 39364, Product Name: Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind, Predicted Profit: 4914.52

Order_ID: 9927, Product Name: Lifetime Advantage Folding Chairs, 4/Carton, Predicted Profit: 3379.05

Order_ID: 30658, Product Name: Sharp AL-1530CS Digital Copier, Predicted Profit: 3345.52

**Bottom 5** Predicted Profit Values:

Order_ID: 36134, Product Name: R280, Predicted Profit: -400.21

Order_ID: 19138, Product Name: Recycled Eldon Regeneration Jumbo File, Predicted Profit: -430.94

Order_ID: 18144, Product Name: Eldon Portable Mobile Manager, Predicted Profit: -463.42

Order_ID: 36646, Product Name: Tennsco Industrial Shelving, Predicted Profit: -686.56

Order_ID: 41696, Product Name: Hoover Portapower Portable Vacuum, Predicted Profit: -1263.36

**Output 4:**

Enter a Region: **Atlantic**

**Top 5** Predicted Profit Values:

Order_ID: 25315, Product Name: Holmes Harmony HEPA Air Purifier for 17 x 20 Room, Predicted Profit: 3450.40

Order_ID: 4416, Product Name: GBC DocuBind P100 Manual Binding Machine, Predicted Profit: 2529.75

Order_ID: 28836, Product Name: 688, Predicted Profit: 2340.30

Order_ID: 7015, Product Name: T65, Predicted Profit: 2253.34

Order_ID: 59234, Product Name: Kensington 7 Outlet MasterPiece Power Center, Predicted Profit: 2059.59

**Bottom 5** Predicted Profit Values:

Order_ID: 8007, Product Name: StarTAC 8000, Predicted Profit: -764.01

Order_ID: 36677, Product Name: Canon Image Class D660 Copier, Predicted Profit: -825.95

Order_ID: 13604, Product Name: Eldon ClusterMat Chair Mat with Cordless Antistatic Protection, Predicted Profit: -1399.32

Order_ID: 53894, Product Name: Lesro Sheffield Collection Coffee Table, End Table, Center Table, Corner Table, Predicted Profit: -1581.44

Order_ID: 32199, Product Name: Canon imageCLASS 2200 Advanced Copier, Predicted Profit: -4656.04

**Output 5:**

Enter a Region: **Prarie**

**Top 5** Predicted Profit Values:

Order_ID: 48614, Product Name: T28 WORLD, Predicted Profit: 1673.58

Order_ID: 24386, Product Name: Gyration Ultra Cordless Optical Suite, Predicted Profit: 1659.33

Order_ID: 54753, Product Name: T39m, Predicted Profit: 1374.11

Order_ID: 45248, Product Name: T28 WORLD, Predicted Profit: 922.78

Order_ID: 50404, Product Name: 24 Capacity Maxi Data Binder Racks, Pearl, Predicted Profit: 913.26

**Bottom 5** Predicted Profit Values:

Order_ID: 50404, Product Name: Lesro Round Back Collection Coffee Table, End Table, Predicted Profit: -366.74

Order_ID: 26016, Product Name: Micro Innovations Media Access Pro Keyboard, Predicted Profit: -372.63

Order_ID: 5538, Product Name: 8290, Predicted Profit: -626.81

Order_ID: 55138, Product Name: Tennsco Industrial Shelving, Predicted Profit: -663.71

Order_ID: 54501, Product Name: Tennsco Commercial Shelving, Predicted Profit: -1138.62

**Output 6:**

Enter a Region: **West**

**Top 5** Predicted Profit Values:

Order_ID: 7427, Product Name: Hoover WindTunnel Plus Canister Vacuum, Predicted Profit: 4276.74

Order_ID: 52035, Product Name: Canon PC1060 Personal Laser Copier, Predicted Profit: 3912.76

Order_ID: 1221, Product Name: GBC DocuBind TL300 Electric Binding System, Predicted Profit: 3674.96

Order_ID: 22657, Product Name: Hayes Optima 56K V.90 Internal Voice Modem, Predicted Profit: 3352.21

Order_ID: 5988, Product Name: Sharp AL-1530CS Digital Copier, Predicted Profit: 3280.86

**Bottom 5** Predicted Profit Values:

Order_ID: 11553, Product Name: Gould Plastics 9-Pocket Panel Bin, 18-3/8w x 5-1/4d x 20-1/2h, Black, Predicted Profit: -516.38

Order_ID: 39301, Product Name: Adesso Programmable 142-Key Keyboard, Predicted Profit: -519.58

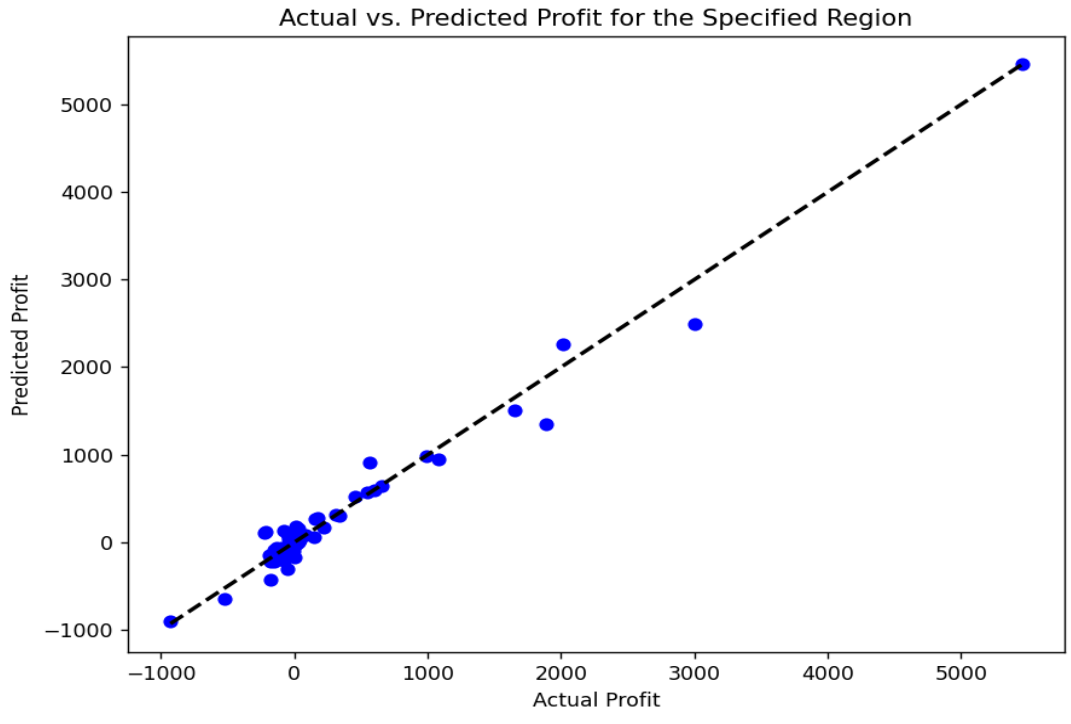
Order_ID: 38758, Product Name: Euro Pro Shark Stick Mini Vacuum, Predicted Profit: -640.94

Order_ID: 22980, Product Name: Eldon Simplefile Box Office , Predicted Profit: -727.95

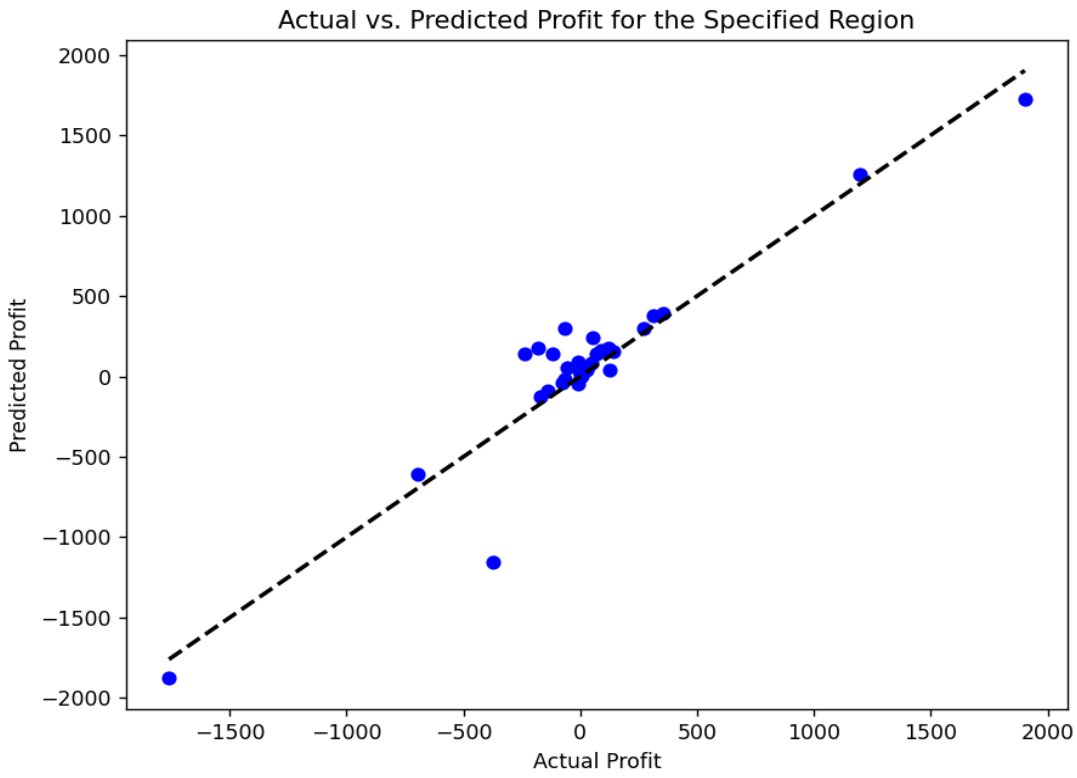
Order_ID: 44320, Product Name: Laminate Occasional Tables, Predicted Profit: -1665.67

**LINEAR REGRESSION VISUALIZATION OF EACH REGION :**

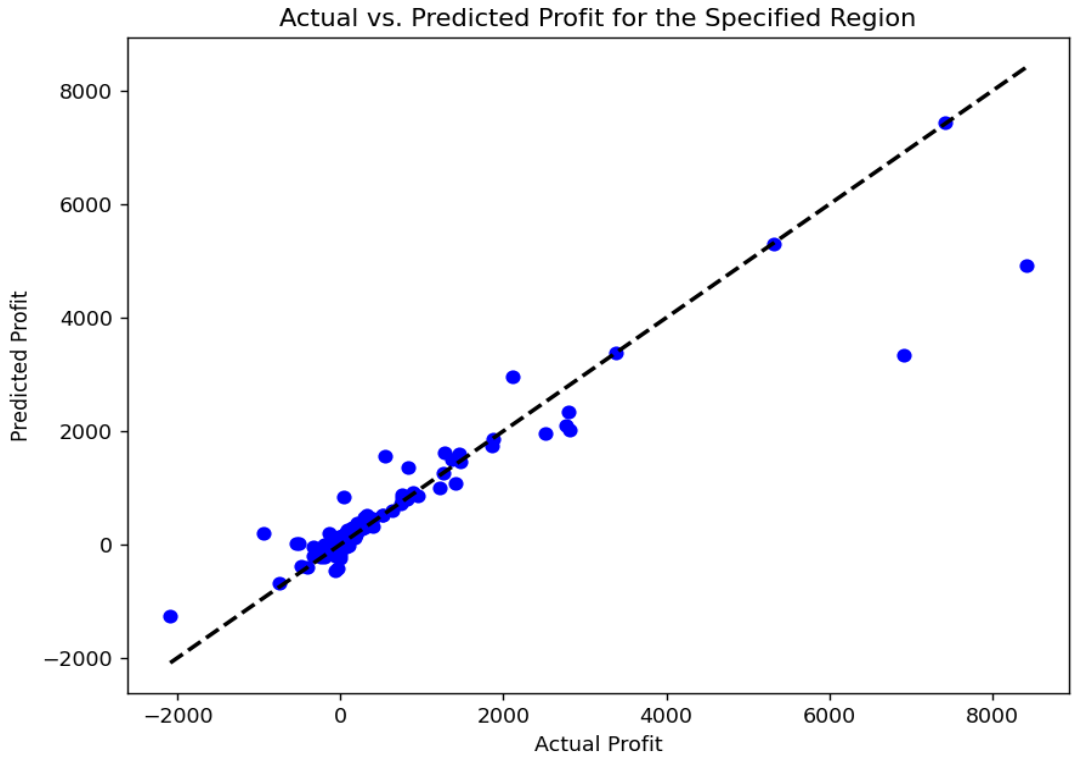
**ONTARIO:**



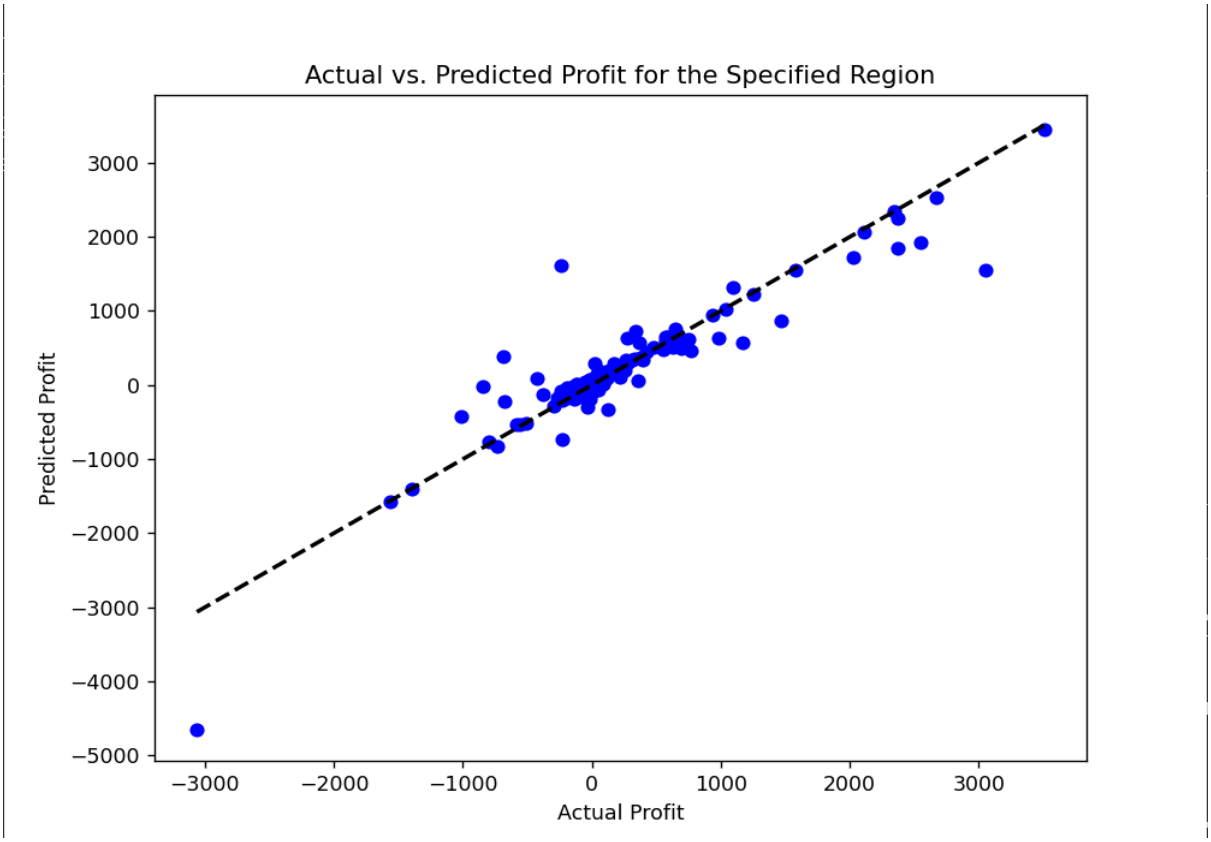
**NUNAVUT:**



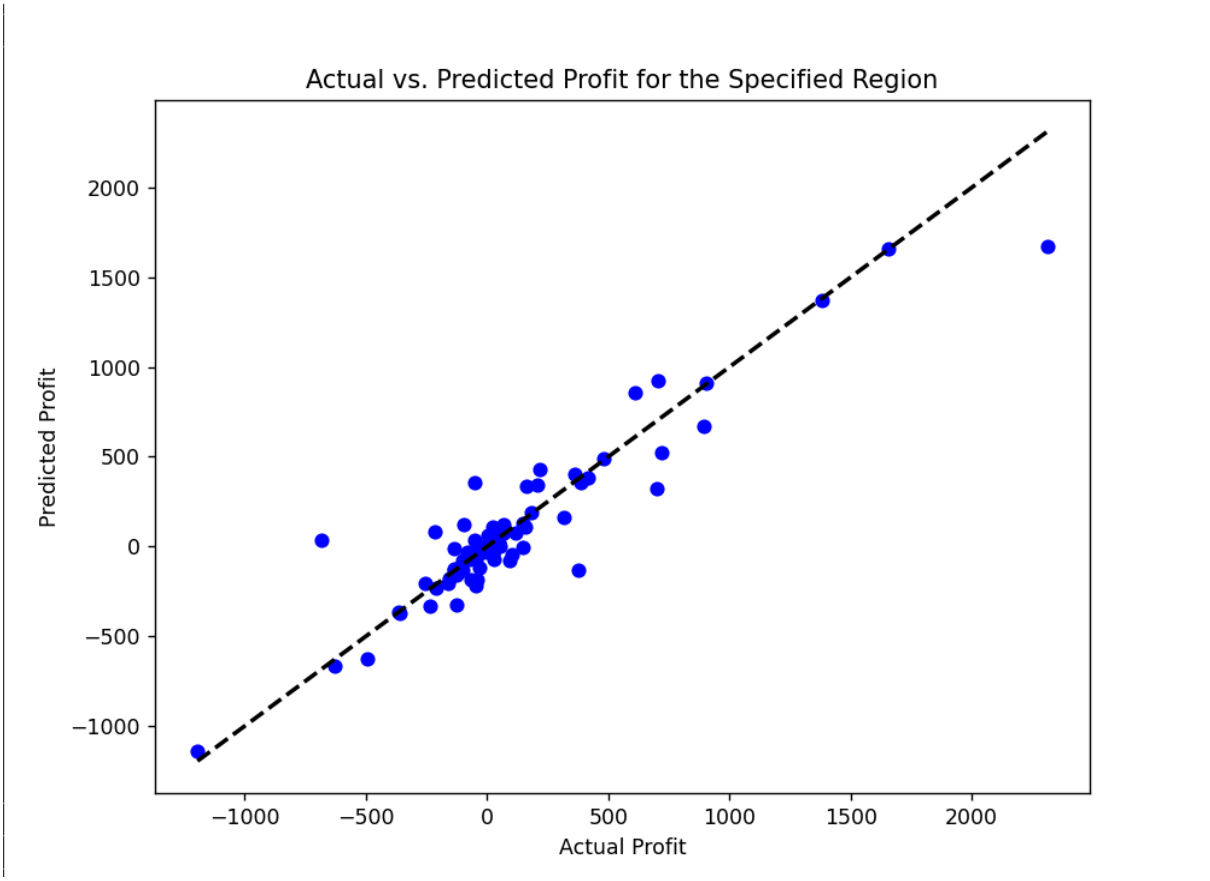

**NORTHWEST TERRITORIES:**



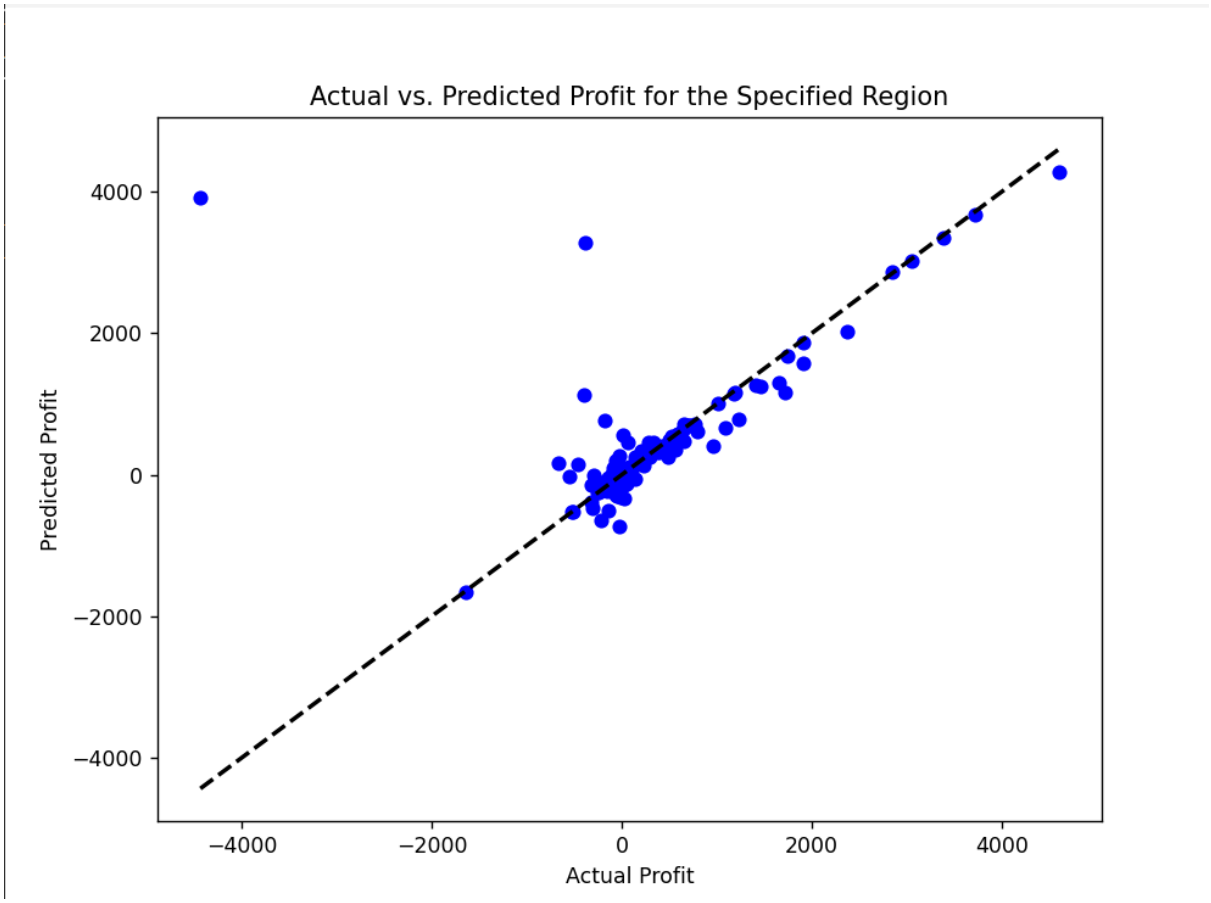
**ATLANTIC:**




**PRARIE:**



**WEST:**




## IBM COGNOS ANALYTICS

### DEFINITION:

IBM Cognos Analytics is a business intelligence and data analytics platform developed by IBM. It is designed to help organizations extract valuable insights from their data, make data-driven decisions, and create interactive reports and dashboards. Here are some key features and aspects of IBM Cognos Analytics.

### MAIN KEY FEATURES IBM COGNOS ANALYTICS:

- **Data Integration:**
  Cognos Analytics provides seamless integration with various data sources, making it easier to access and analyze data from different platforms.
- **Report Authoring:**
  Users can create interactive and customizable reports using a user-friendly web interface, simplifying the process of report creation.
- **Dashboard Creation:**
  The platform enables the development of interactive dashboards that combine multiple visualizations for a holistic view of data.
- **Ad Hoc Reporting:**
  Cognos Analytics supports ad-hoc reporting, allowing users to generate reports on-the-fly without predefined templates.

- **Advanced Analytics:**
  It offers advanced analytics and predictive analytics capabilities, allowing organizations to build and deploy machine learning models for data-driven insights and predictions.

## ABOUT PRODUCT SALES ANALYSIS IN IBM COGNOS ANALYTICS:

IBM Cognos Analytics provides a range of powerful tools for visualizing product sales analysis datasets. It allows you to create interactive and insightful visualizations that can help you uncover patterns, trends, and key insights from your sales data. Here's how you can visualize a product sales analysis dataset in IBM Cognos Analytics:

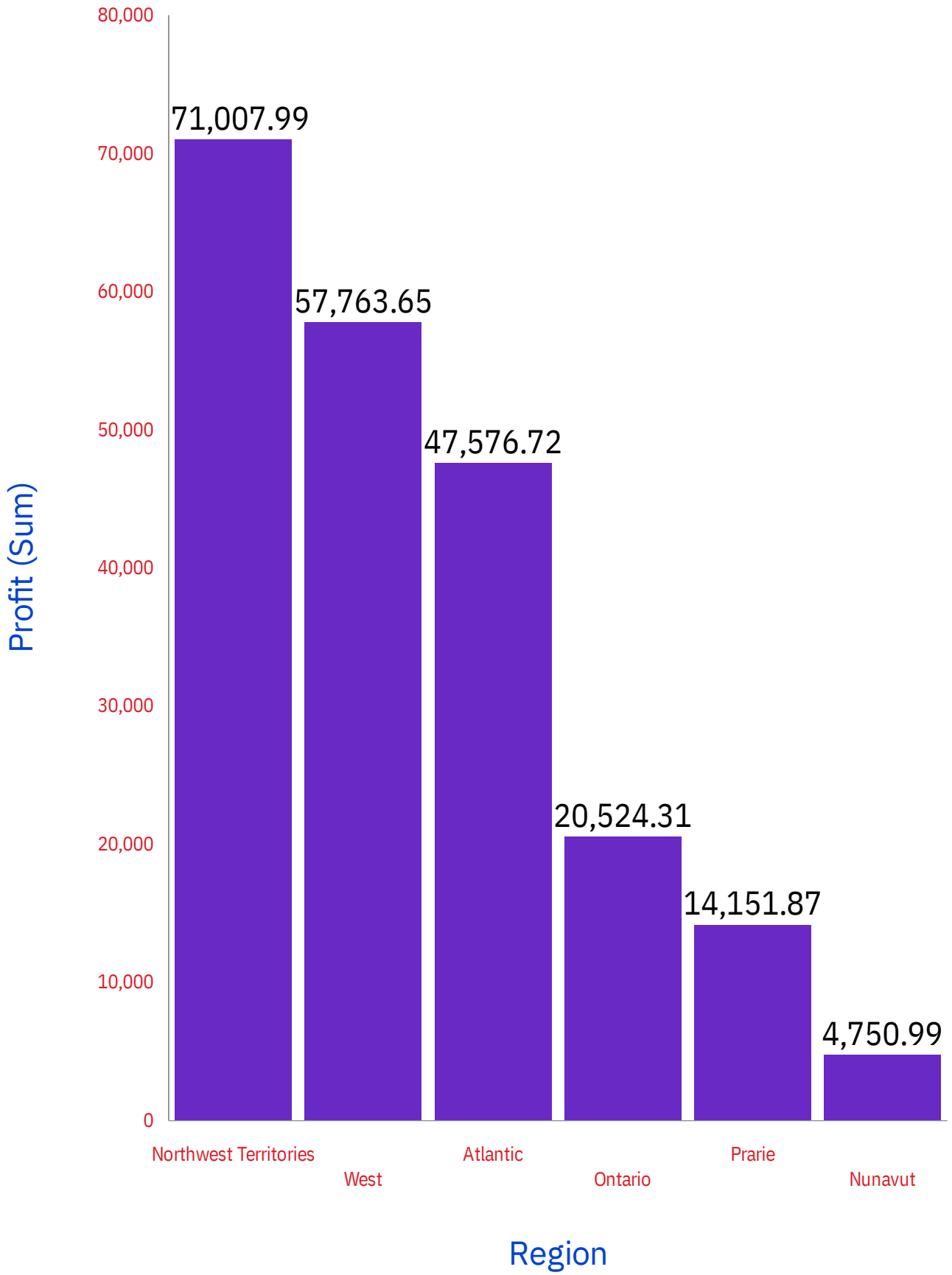
- **Data Connection:**
  Start by connecting your product sales dataset to IBM Cognos Analytics. This can be done by importing your data into Cognos or connecting directly to your data source.
- **Data Preparation:**
  Before visualization, ensure your data is clean and structured. You may need to perform data cleaning and transformation tasks within Cognos Analytics.
- **Create Data Models:**
  Build data models within Cognos Analytics to define the relationships, hierarchies, and calculations necessary for your analysis.
- **Common Visualizations:**
  Cognos Analytics offers a variety of common visualization options, including bar charts, line charts, pie charts, scatter plots, and more. You can easily select and configure these visualizations to represent your sales data.
- **Interactive Dashboards:**
  Assemble multiple visualizations into interactive dashboards. These dashboards can display product sales data from different perspectives, allowing users to explore and interact with the data.
- **Filters and Parameters:**
  Implement filters and parameters in your visualizations to allow users to dynamically adjust the data they're viewing. For instance, users can select specific time periods or product categories for analysis.
- **Trend Analysis:**
  Use line charts to show sales trends over time. You can configure these charts to include multiple product lines for easy comparison.
- **Comparative Analysis:**
  Create visualizations that compare the performance of different product categories, regions, or sales channels. Stacked bar charts, scatter plots, and heatmaps can be useful for this purpose.

## CATEGORIES TO BE CONSIDERED FOR OUR VISUALIZATION :

- ✓ **Order_ID :** Top 5 and Bottom 5 Order_ID's of each region.
- ✓ **Product_Name :** Top 5 and Bottom 5 Product_Name 's of each region.
- ✓ **Profit :** Top 5 (positive) and Bottom 5 (negative) profits of each region.
- ✓ **Region :** There are five regions in the dataset. They are **Atlantic , Ontario , Prarie , Northwest Territories , Nunavut , West .**

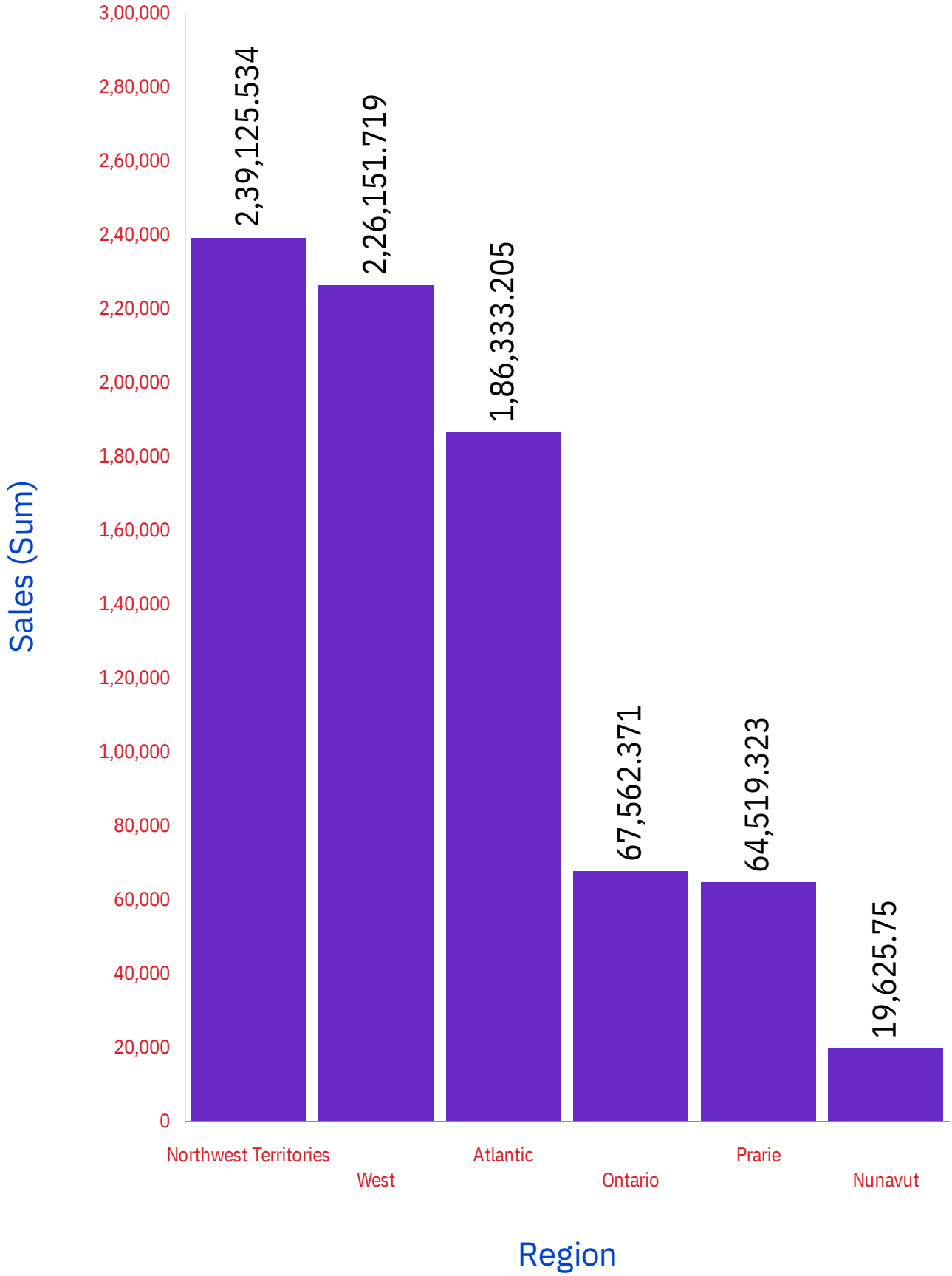
**By using these parameters on region wise visualization we can clearly understand that which top 5 products produced more profit and which bottom 5 products produced less profit.**

PROFIT BY REGION



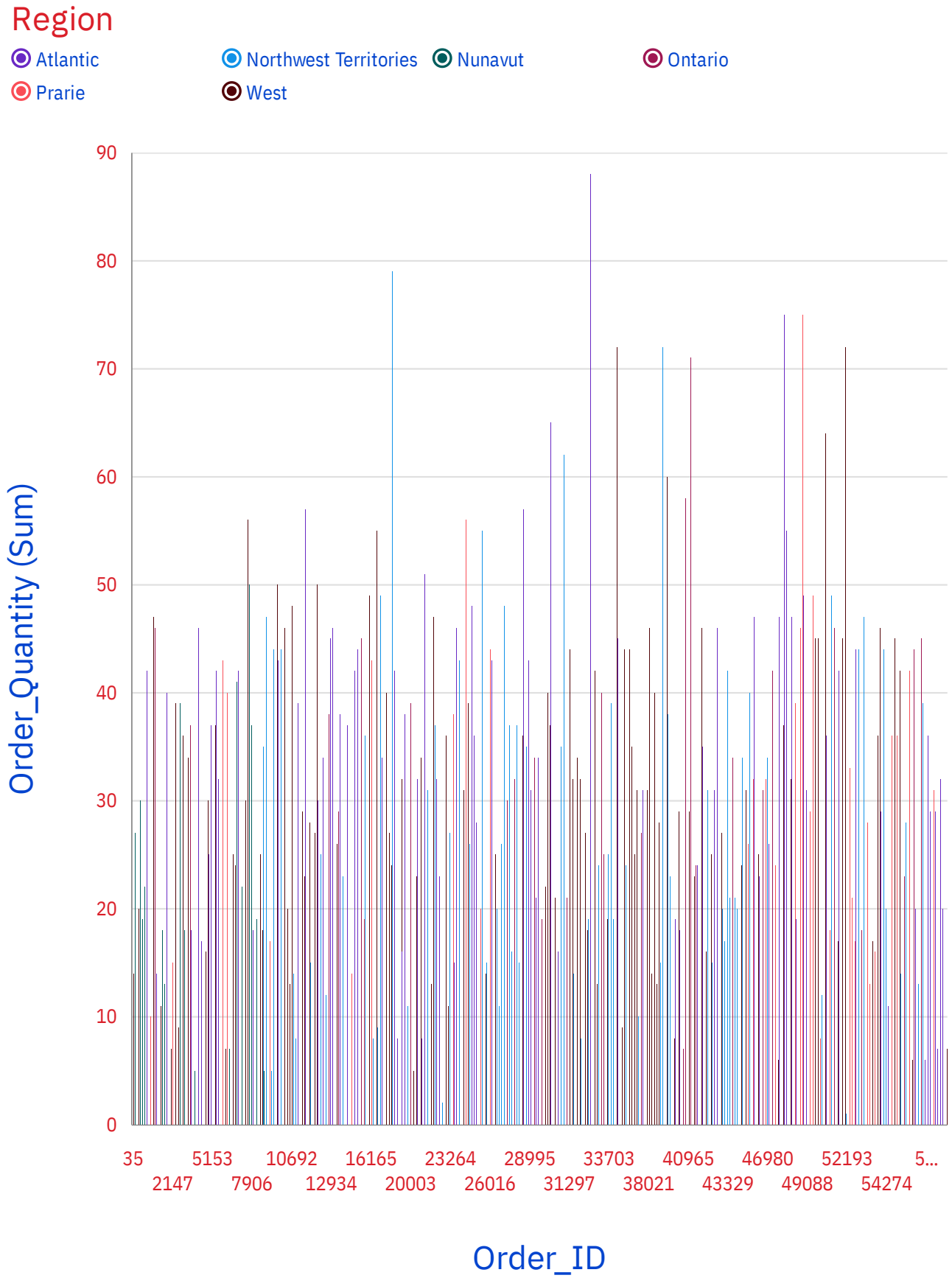
SALES BY REGION

## *Sales by Region*



ORDER ID &QUANTITY

## *Order_Quantity by Order_ID colored by Region*



C_SEG Vs REGION

## *Customer_Segment colored by Region*

Region

- Atlantic
- Northwest Territories
- Nunavut
- Ontario
- Prarie
- West

REGION Vs PR_CATEGORY

## *Product_Category colored by Region*

Region

Atlantic
Northwest Territories
Nunavut
Ontario
Prarie
West

REG Vs PR_SUBCATEGORY

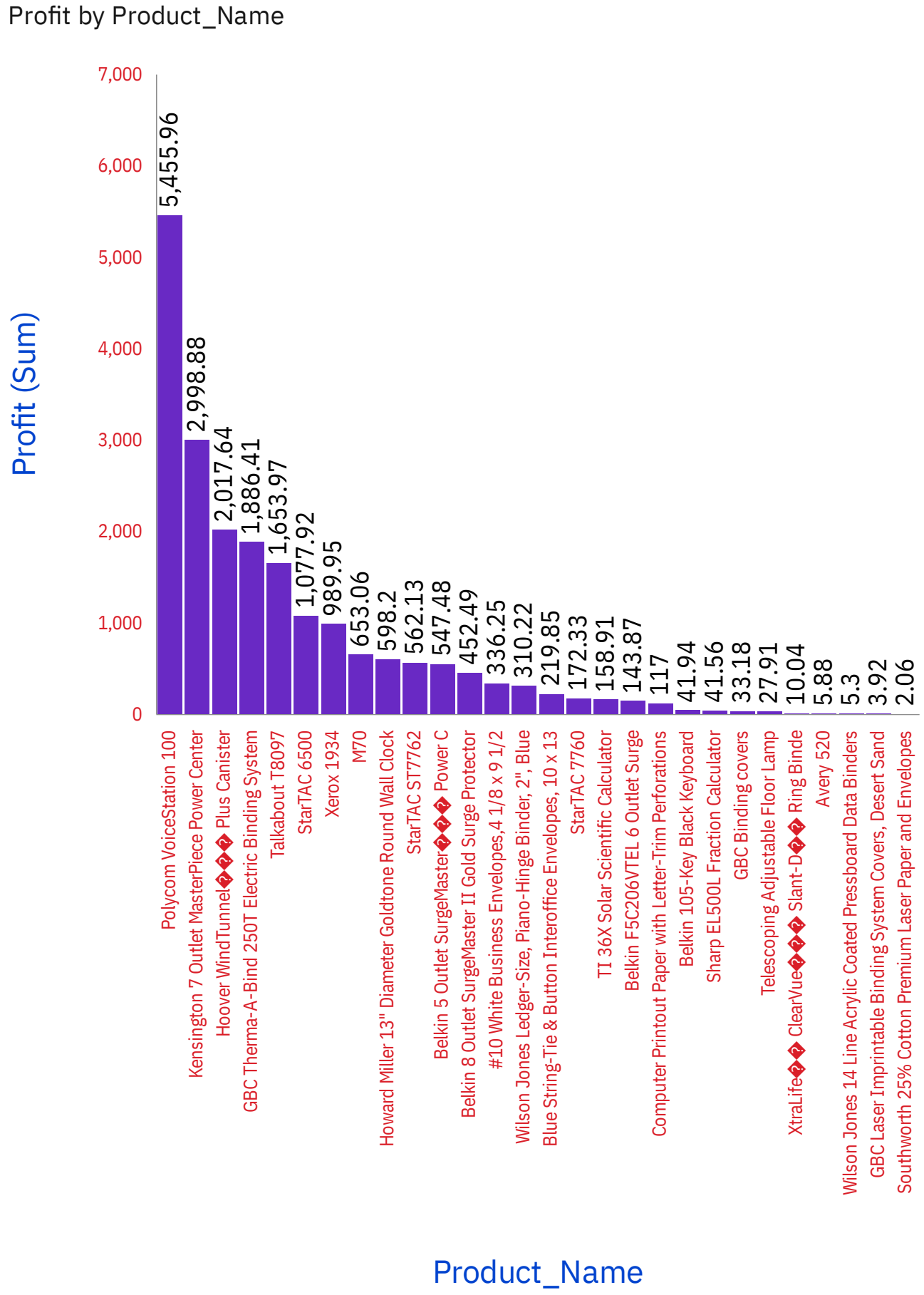
## *Product_Sub-Category colored by Region*

Region

Atlantic Northwest Territories Nunavut Ontario

Prarie West

REGION WISE PRODUCT PROFIT



REGION Vs PRODUCT_CONTAINER

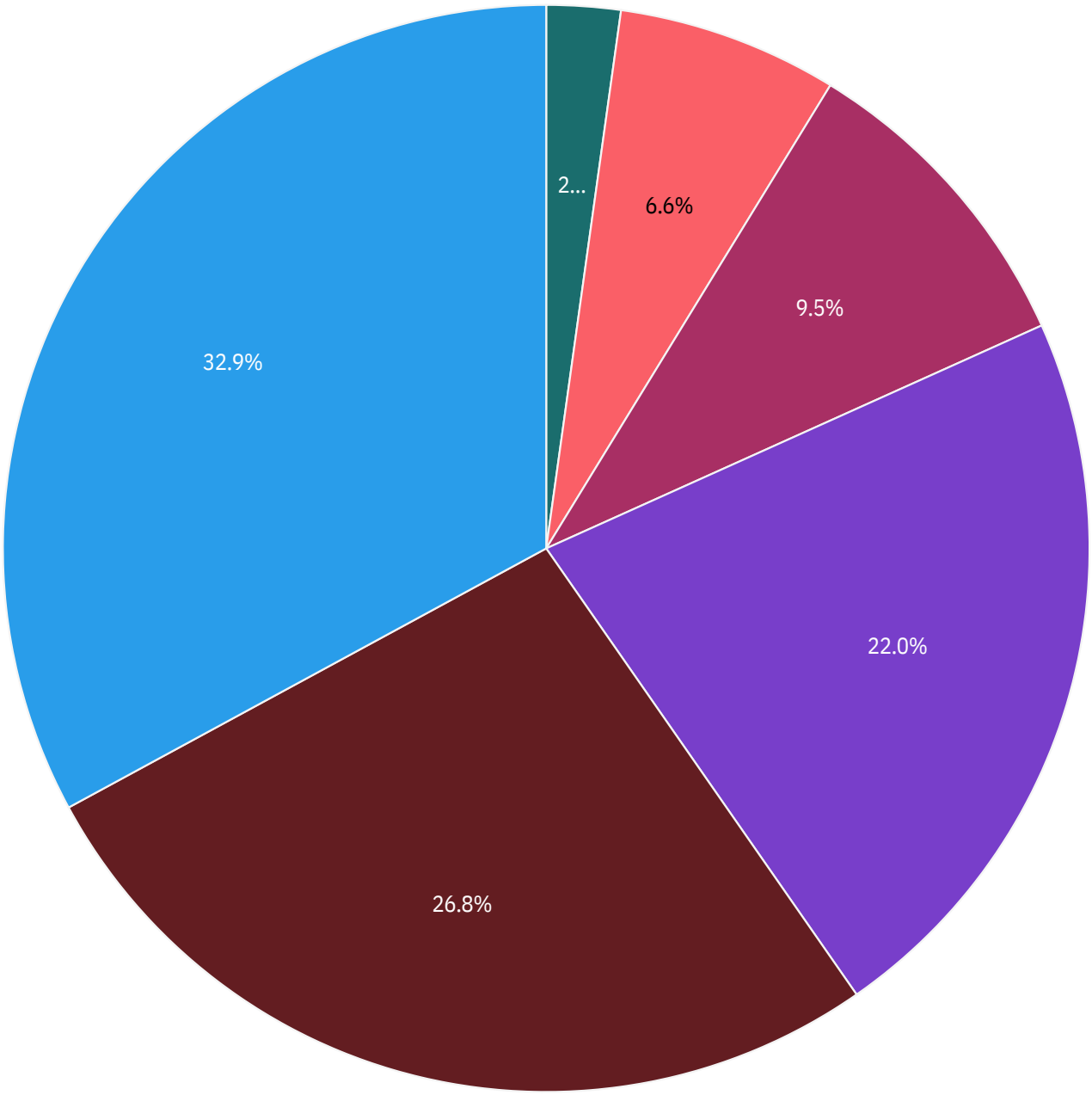
## *Profit by Region*

Region

Nunavut Prarie Ontario Atlantic

West Northwest Territories

| Product_Container |
| --- |
| Large Box |
| Medium Box |
| Small Box |

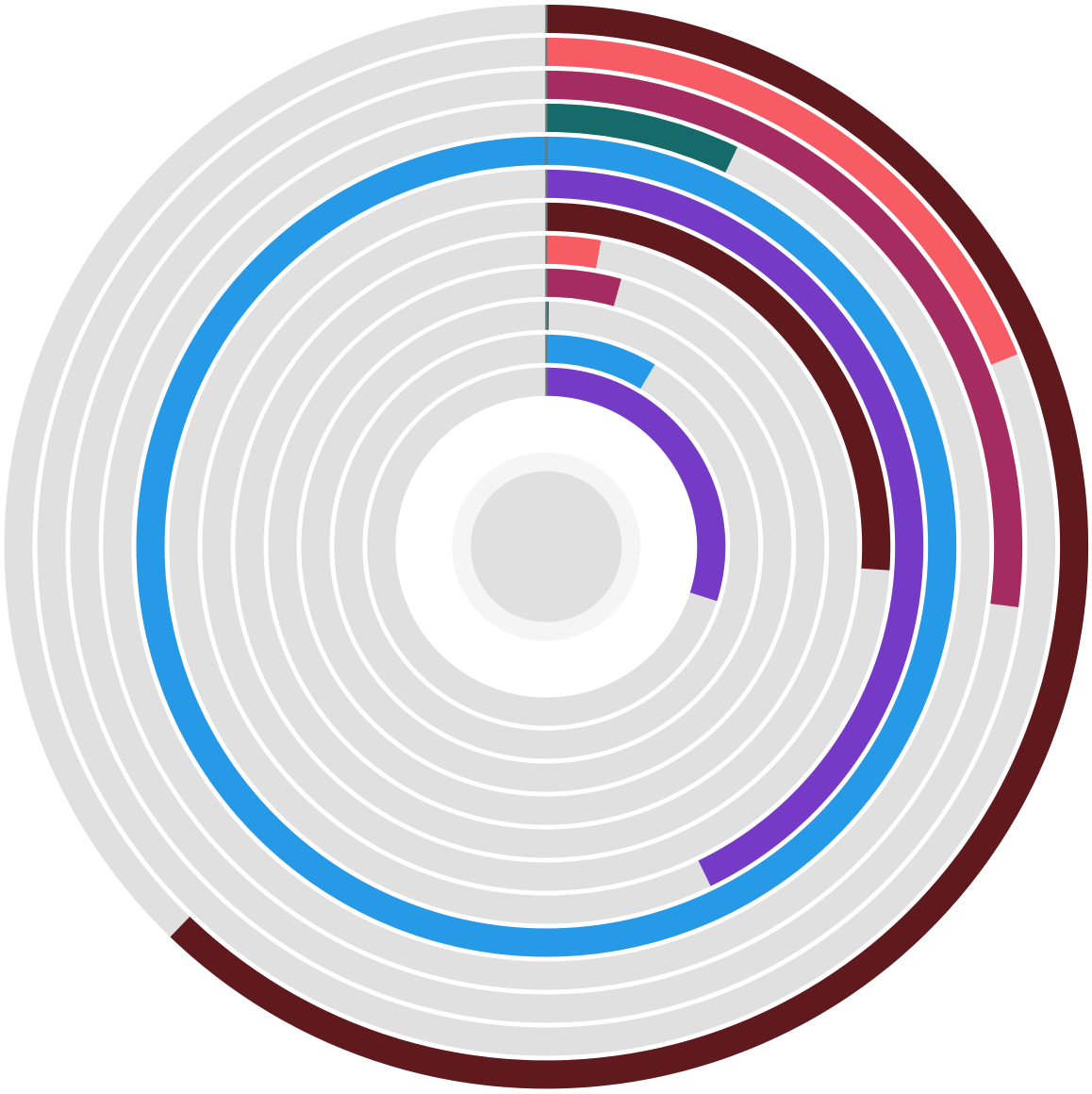


REGION Vs SHIP_MODE

## *Profit by Ship_Mode colored by Region*

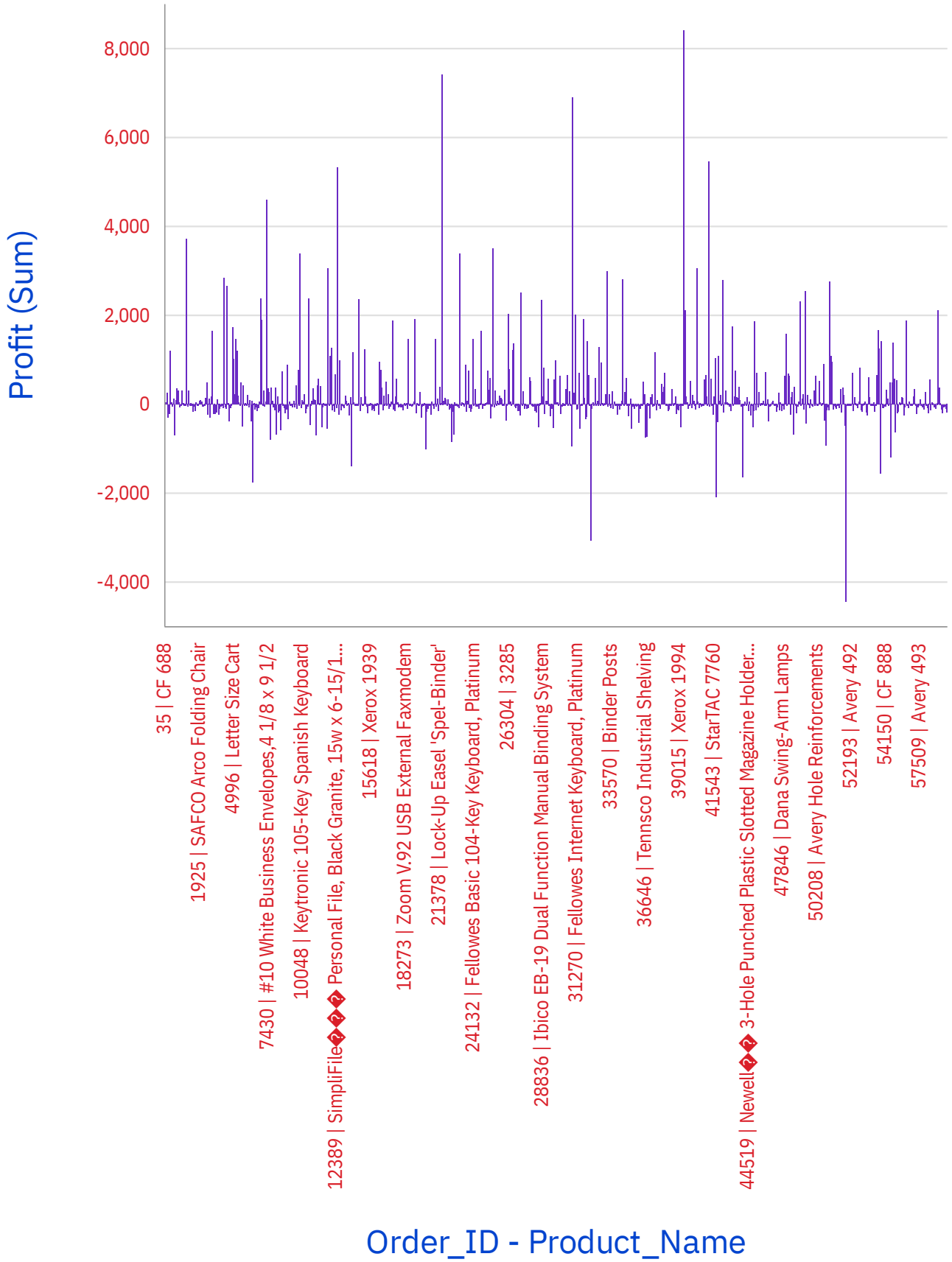
Region

Atlantic
Northwest Territories
Nunavut
Ontario
Prarie
West



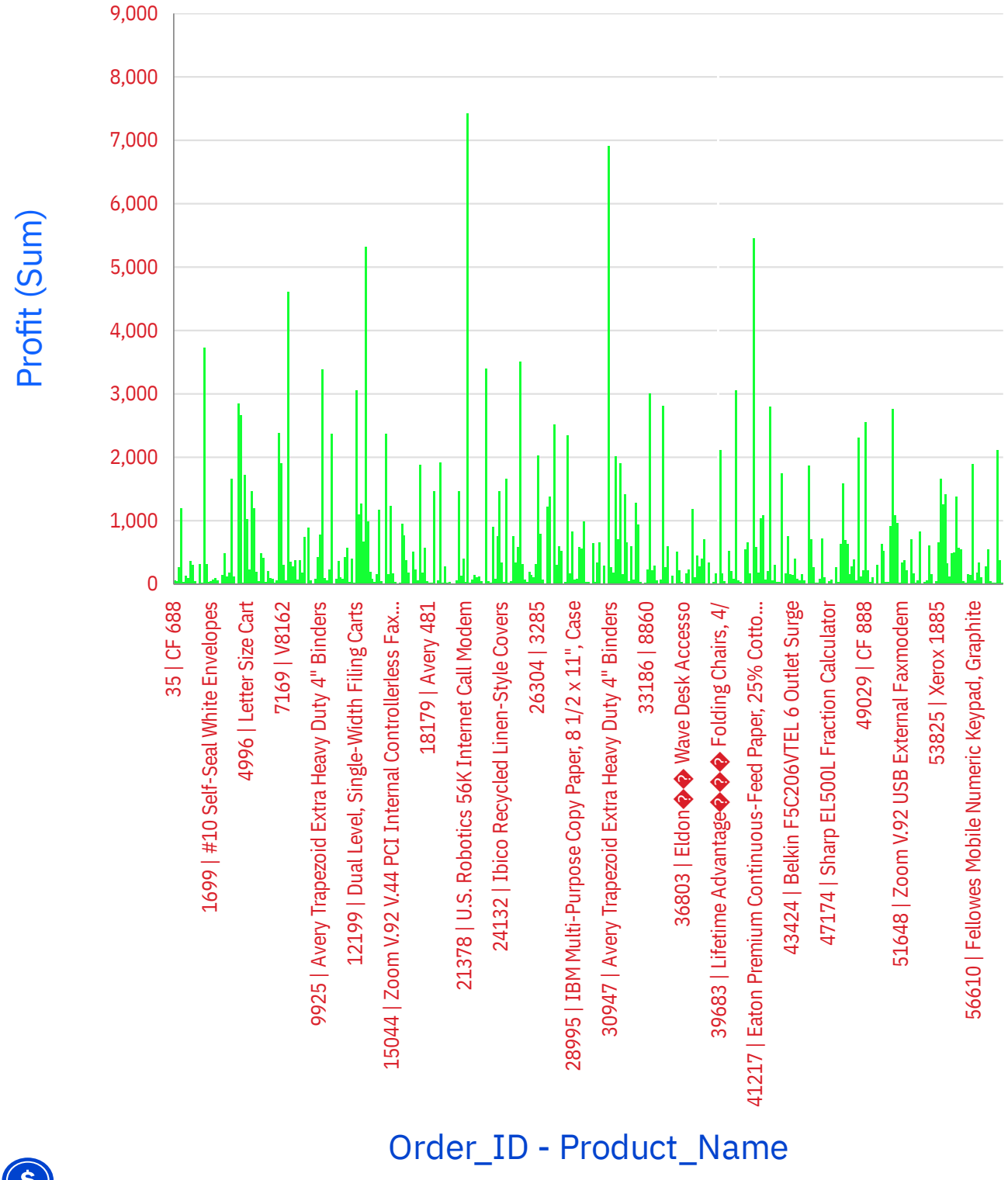
PROFIT

## *Profit by Order_ID and Product_Name*



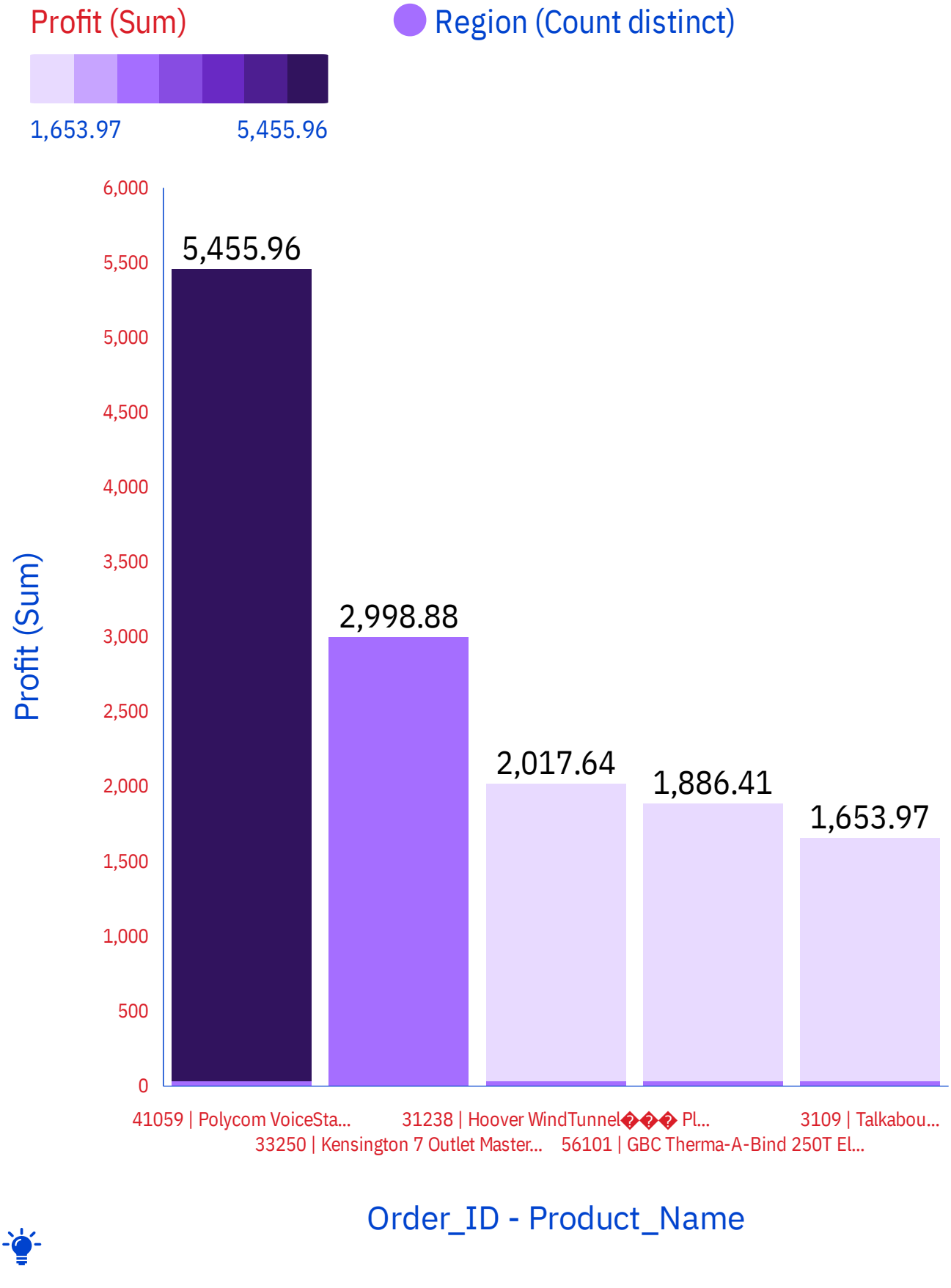
POSITIVE PROFIT

## *Profit by Order_ID and Product_Name colored by Profit*



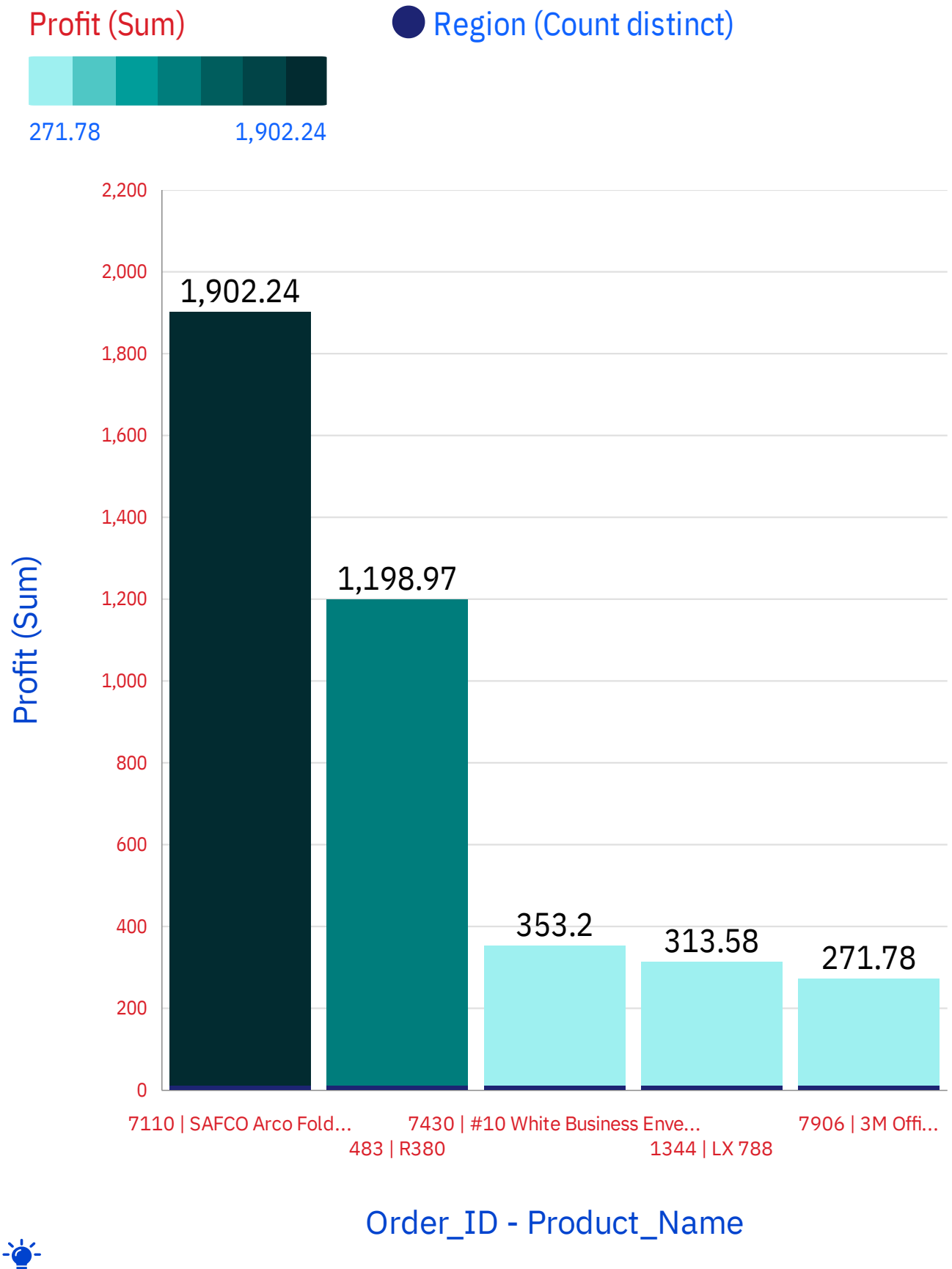
ONTARIO

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



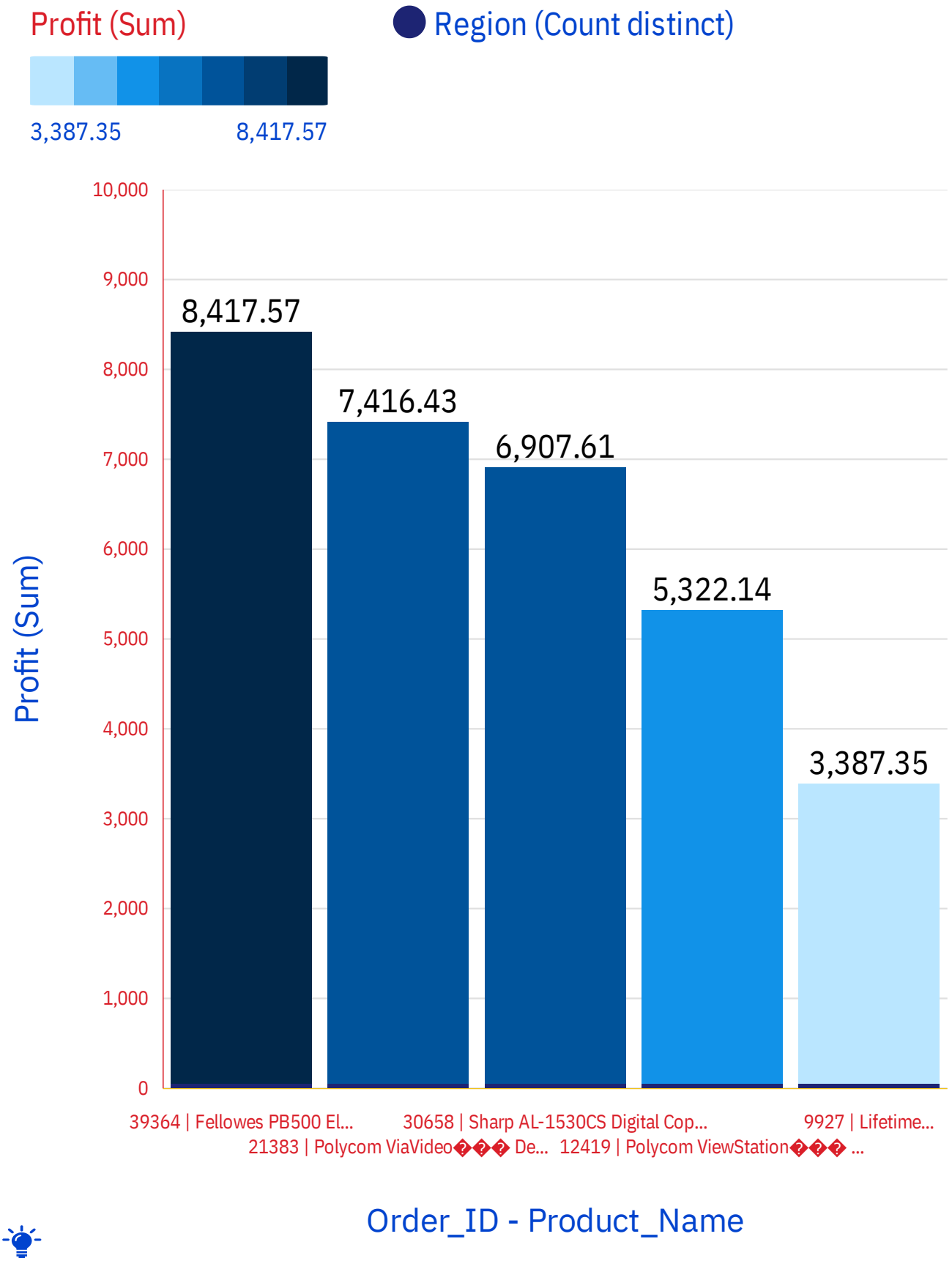
NUNAVUT

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



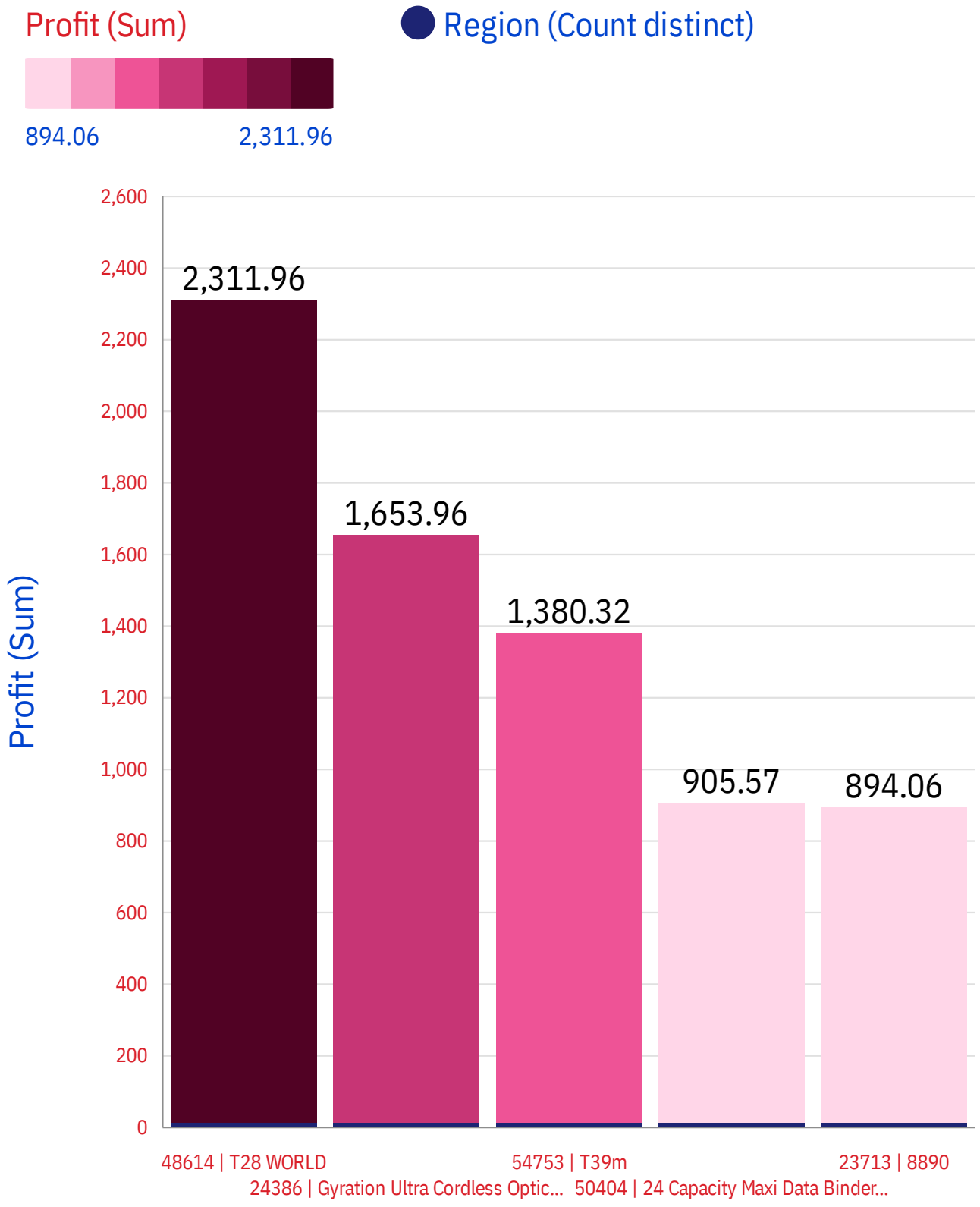
NW TERRITORIES

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



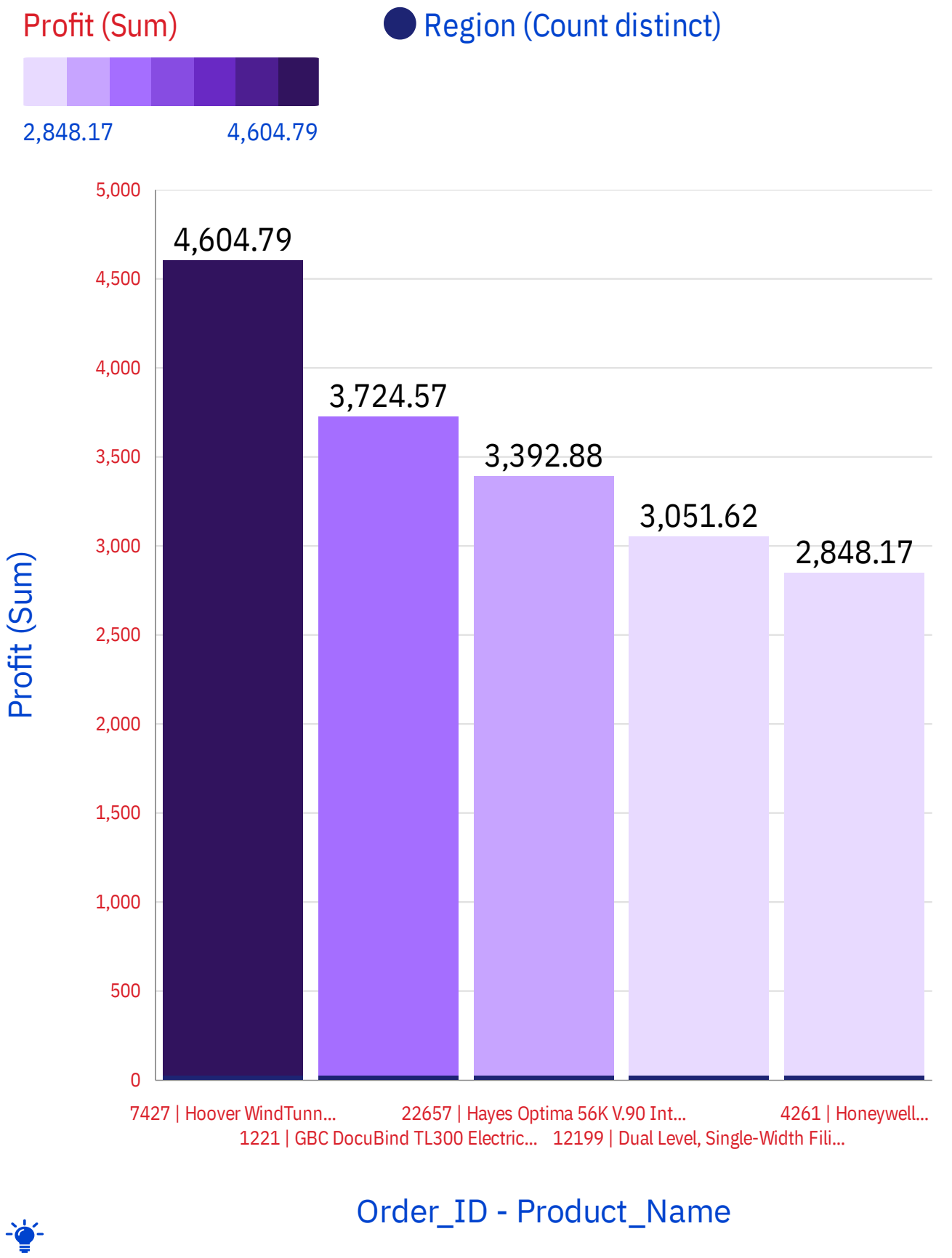
PRARIE

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



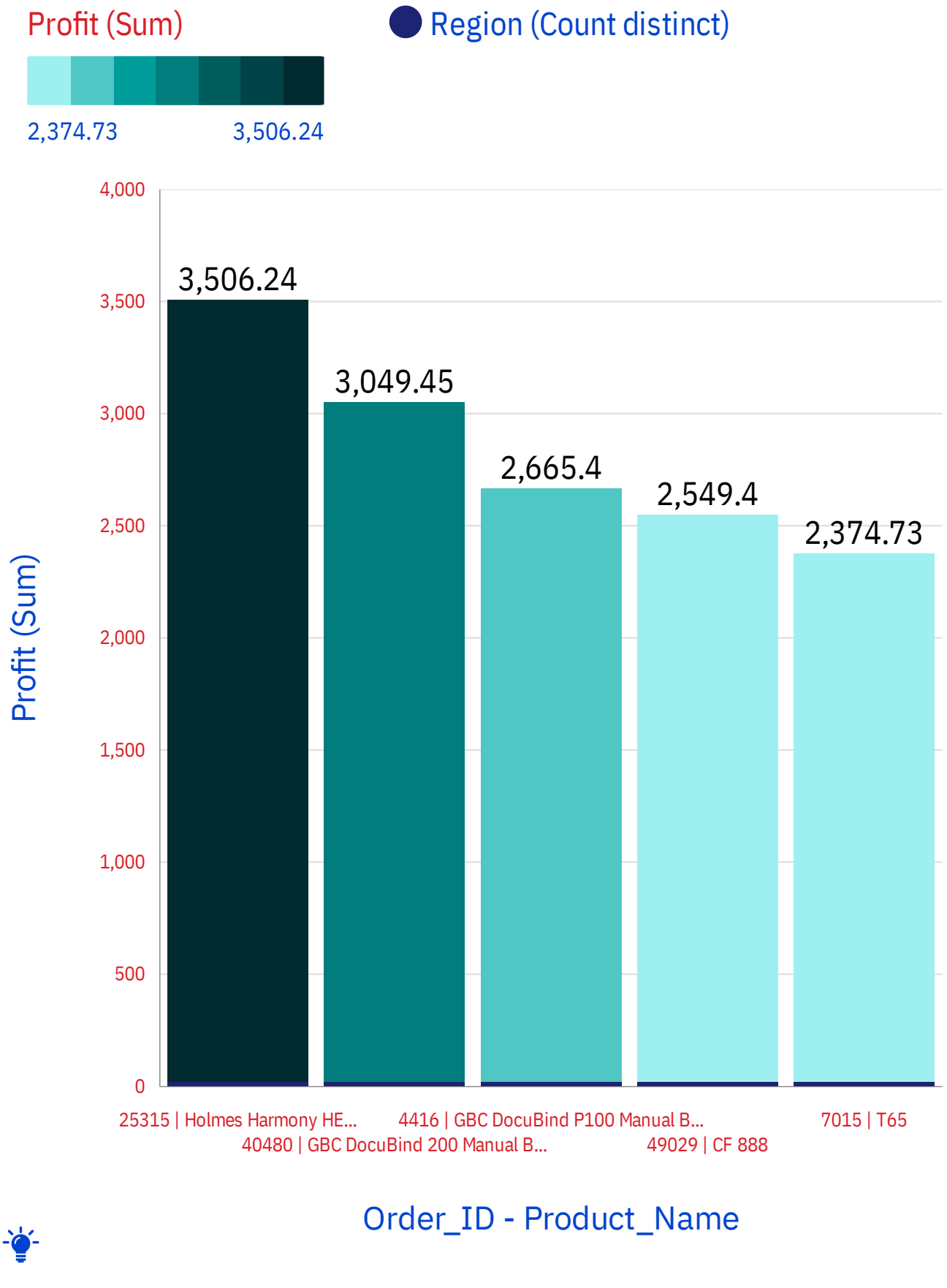
WEST

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



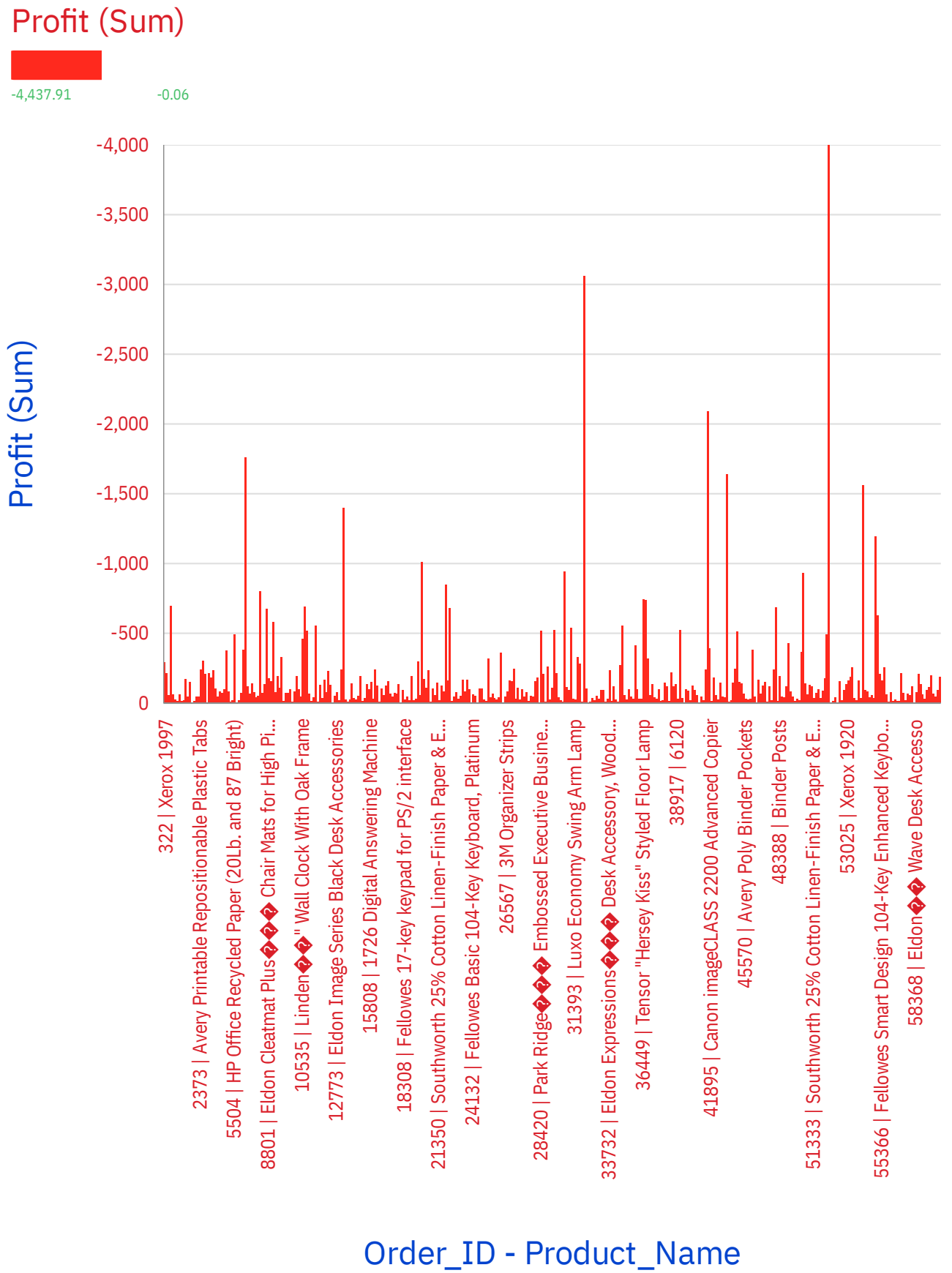
ATLANTIC

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



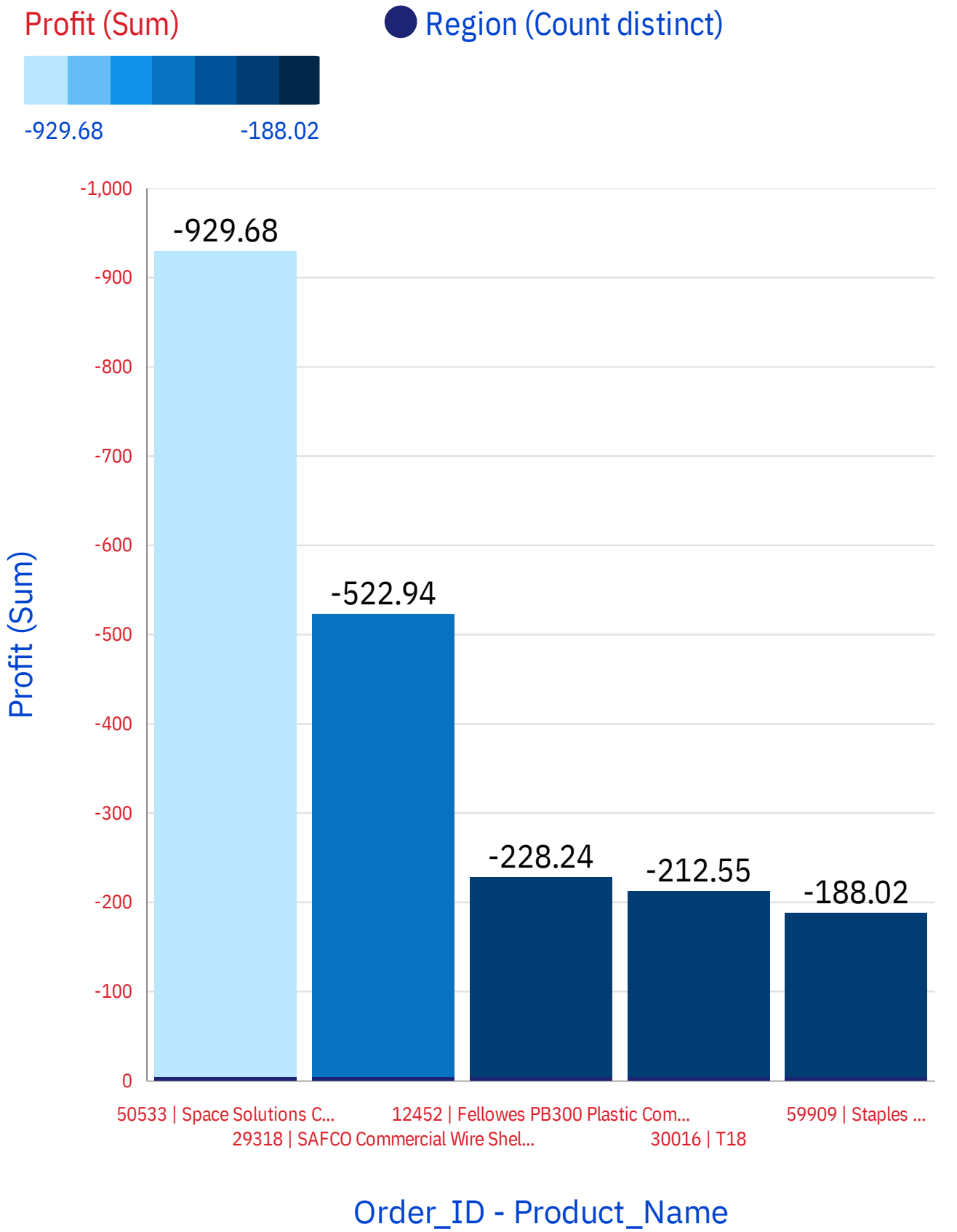
NEGATIVE PROFIT

## *Profit by Order_ID and Product_Name colored by Profit*



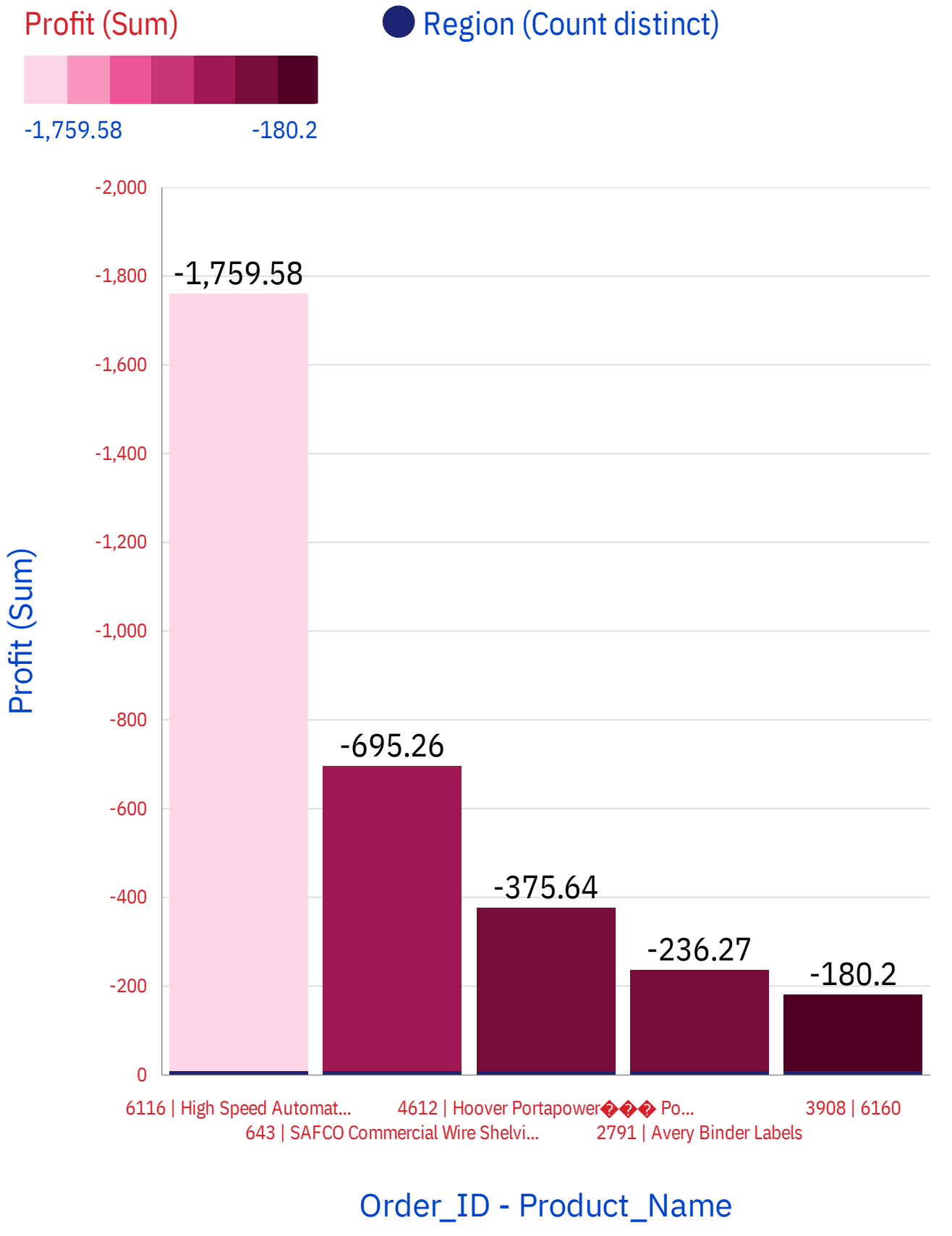
ONTARIO

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



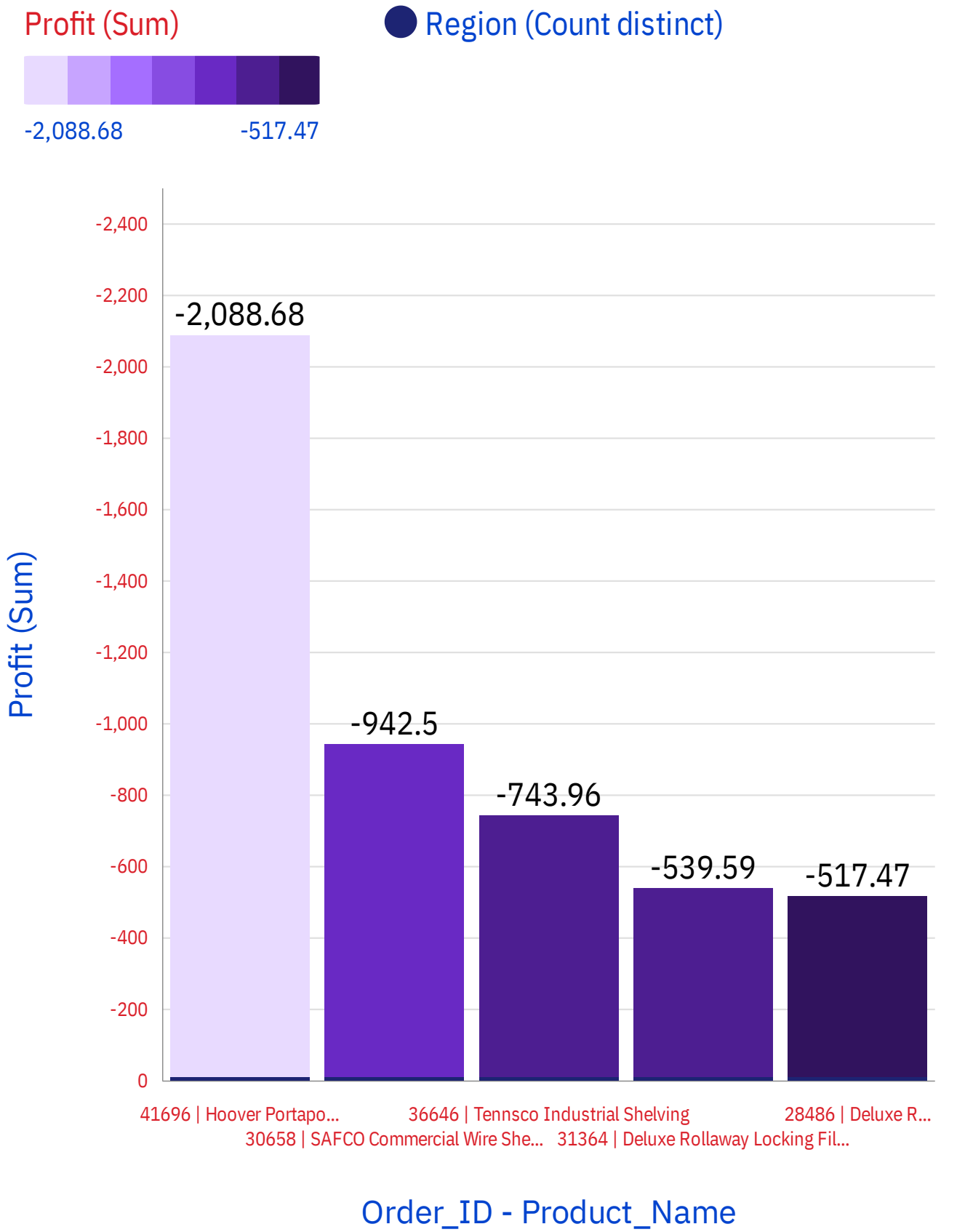
NUNAVUT

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



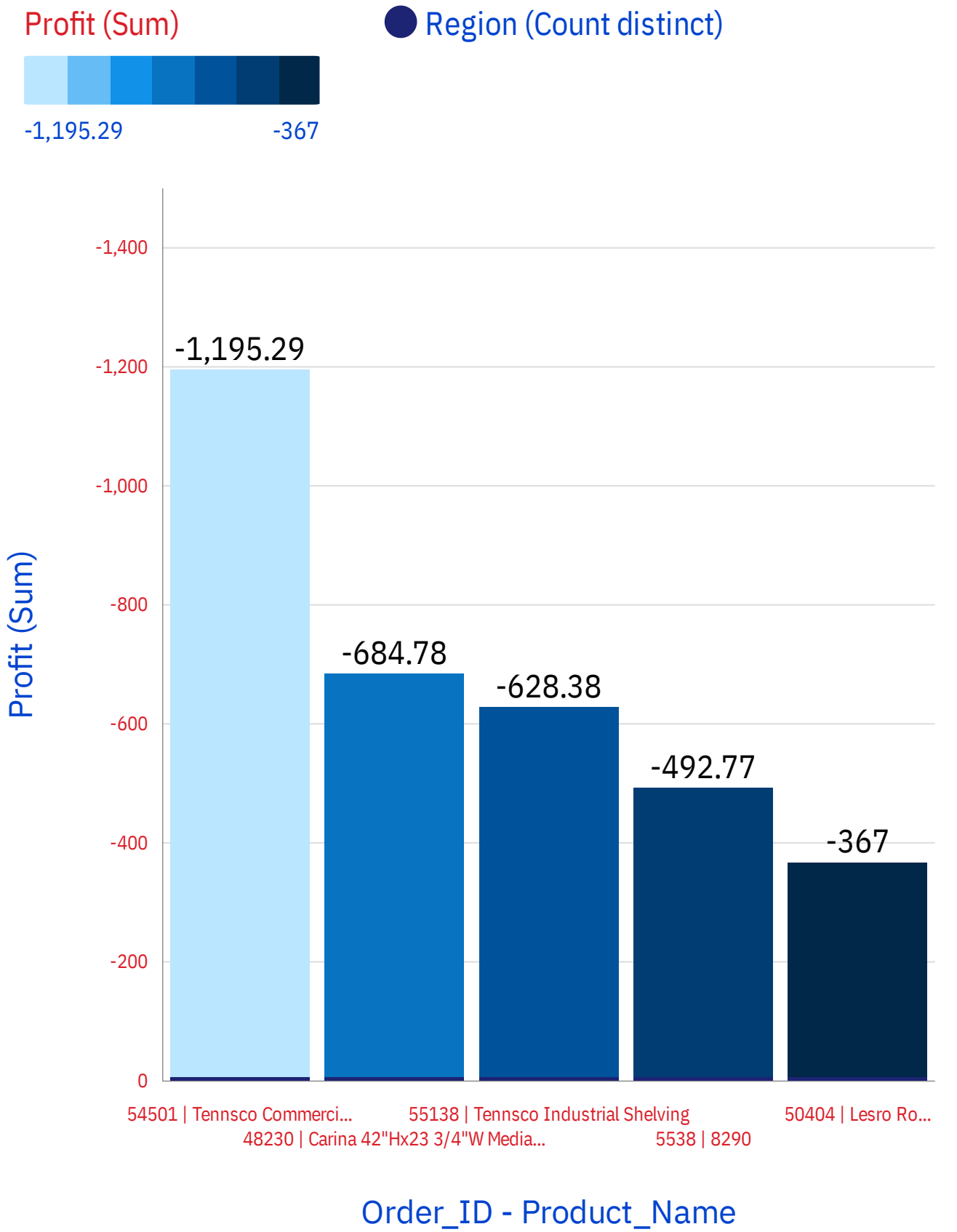
NW TERRITORIES

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



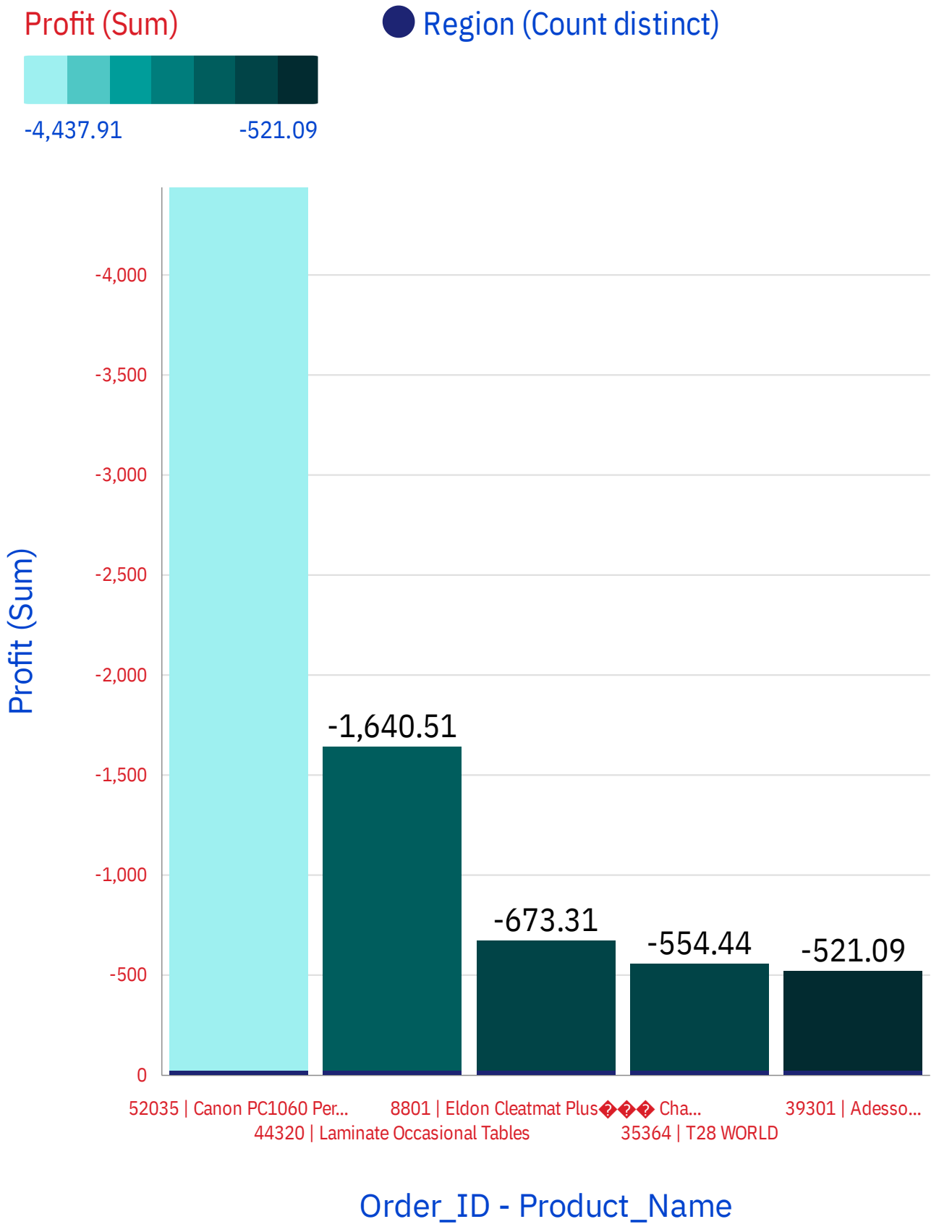
PRARIE

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



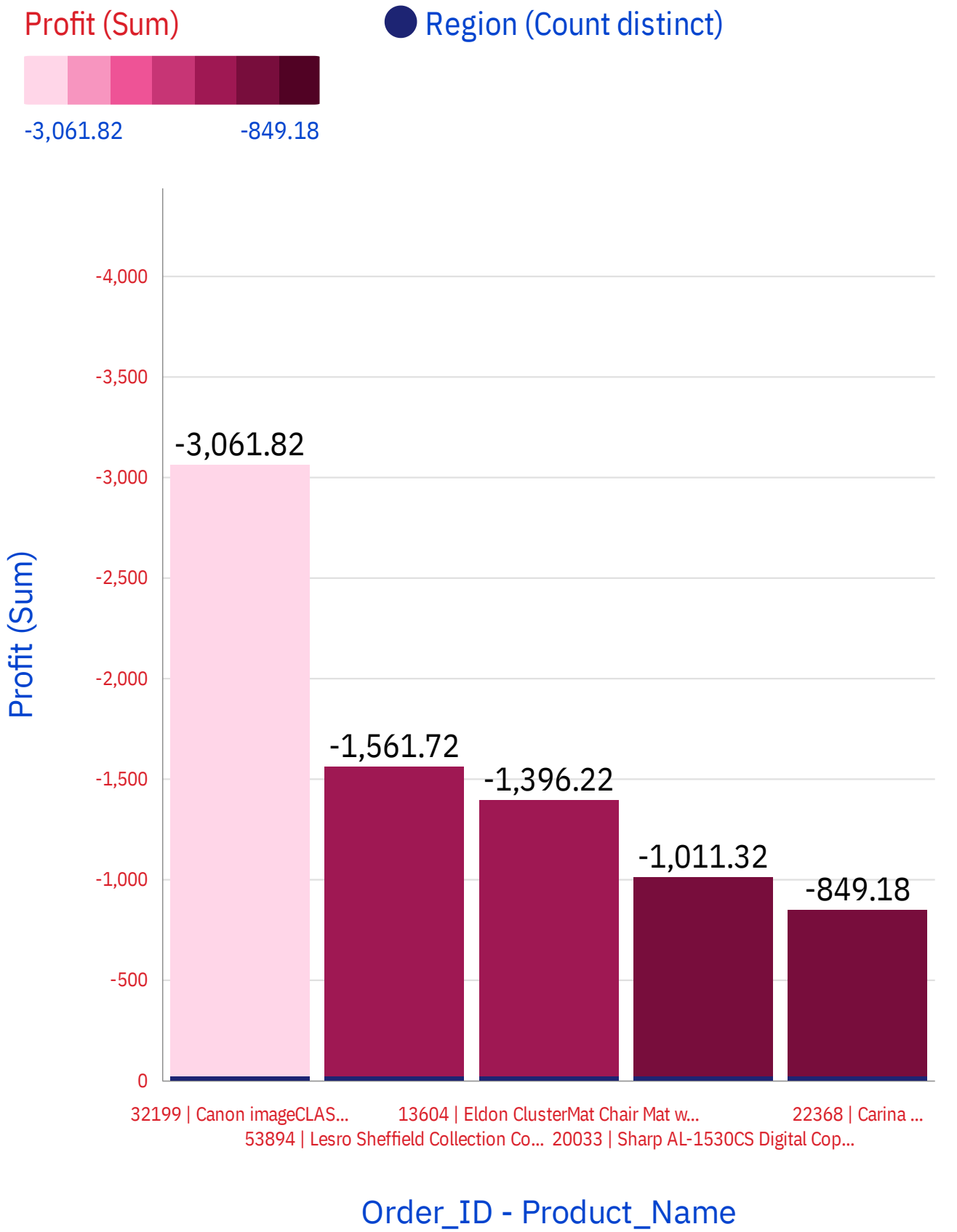
WEST

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



ATLANTIC

## *Region compared to Profit by Order_ID and Product_Name colored by Profit*



## CONCLUSION:

Thus using the **IBM COGNOS ANALYTICS** , the visualization of product sales analysis by the given dataset was done region wise . By using these visualization we can clearly understand that which **top 5** products produces more profit and which **bottom 5** products produces less profit in each region.