

Tariff Fashion Project – Final Report

1. Project Overview

The **Tariff Fashion Project** is a collaborative data science initiative focused on understanding how U.S government-imposed tariffs influence the final retail pricing of fashion products. By leveraging data cleaning techniques, exploratory data analysis (EDA), machine learning models, and interactive visualizations, the team developed a comprehensive system to analyze and predict post-tariff product prices.

The project aimed to:

- Identify how different fashion categories and brands are impacted by tariffs.
- Predict final product prices using key features such as base cost, region, material, and rating.
- Provide an intuitive dashboard and prediction app to interact with the dataset and model.

2. Dataset Summary

The dataset was manually collected and curated by the team, consisting of **1000 fashion products** with the following attributes:

- **Product Category** – e.g. Jeans, T-Shirts, Sneakers
- **Brand** – Brand label of the item
- **Material** – Fabric or material type (e.g. Cotton, Denim)
- **Region** – Sales region or distribution location
- **Base Cost** – Raw cost before profit or tariffs
- **Weight** – Weight of the product (kg)
- **Rating** – User rating score (out of 5)
- **Price Before Tariff** – Selling price before tariff application

- **Price After Tariff** – Price after tariff implementation
- **Tariff Increase Percentage** (*Engineered*) – % increase in price due to tariff

3. Step-by-Step Workflow

Step 1: Dataset Collection & Cleaning

Owner: *VarunRaj & Anand Yadav*

- Assembled a dataset with over 1000 unique fashion product entries
- Cleaned and formatted columns using Pandas
- Removed irrelevant categories (electronics, groceries)
- Handled missing data across columns (e.g., weight, rating)
- Created new column:
$$\text{Tariff Increase \%} = ((\text{Price After Tariff} - \text{Price Before Tariff}) / \text{Price Before Tariff}) * 100$$

Step 2: Exploratory Data Analysis (EDA)

Owner: *Krish Gupta*

- Explored product category vs. price distribution
- Visualized outliers using box plots
- Identified high tariff impact zones by brand and material
- Used Seaborn and Matplotlib to build charts showing pricing shifts across categories
- Observed that low base cost products saw higher tariff percentage jumps

Step 3: Power BI Visualization

Owner: *Smriti Pandey*

- Imported cleaned dataset into Power BI
- Created interactive dashboards featuring:
 - Pre- vs Post-Tariff Price comparisons
 - Tariff Increase % per brand
 - Price delta per category
 - Region-wise insights
- Compiled all visuals into a structured PDF for presentation

Step 4: Machine Learning Model

Owner: *Shraddha Tiwari*

- Built a `RandomForestRegressor` model using `scikit-learn`
- Encoded categorical data using `OneHotEncoder`
- Combined preprocessing and training into a `Pipeline`
- Split dataset 80:20 for training and testing
- Used `RMSE` as performance metric
- Achieved stable prediction performance and interpretability

Step 5: Streamlit App Development

Owner: *Shraddha Tiwari*

- Created an intuitive frontend using **Streamlit**
- Users can input product features (brand, cost, weight, etc.)
- Model predicts post-tariff price live in the browser
- App integrates trained model via `jobjlib`

- Clean UI + responsiveness made it demo-ready for stakeholders

Step 6: Alternate ML Notebook Development

Owner: *Roshni Sharma*

- Built an alternate regression notebook in **Google Colab**
- Used **LinearRegression** to predict post-tariff pricing
- Implemented basic data visualizations and feature selection
- Demonstrated linear relationships between cost-based features and final price
-
-

Step 7: Project QA, Verification & Coordination

Owner: *VarunRaj*

- Collected and curated a dataset of 1000+ fashion products from various sources
- Cleaned the raw data and contributed to removing irrelevant entries
- Verifying final outputs of the fashion tariff project.
- Cross-checked multiple submissions (CSV, notebooks, dashboards) before compiling the final report.

4. Algorithms & Tools Used

Category

Tools/Tech

Data Handling

Pandas, NumPy

Visualization	Matplotlib, Seaborn, Power BI
ML Model	RandomForestRegressor, LinearRegression
Preprocessing	OneHotEncoder, ColumnTransformer
Evaluation	Root Mean Squared Error (RMSE)
Web App	Streamlit
Collaboration	Google Colab, Jupyter Notebook, ,Notion

5. Final Observations & Conclusion

The **Tariff Fashion Project** successfully illustrated how tariffs affect fashion pricing and how machine learning can be applied to predict these effects with reasonable accuracy. Power BI dashboards and visualizations offered a digestible way to observe pricing trends and tariff impact. The Streamlit app gave life to the model through real-time predictions, while multiple modeling approaches gave depth to our experimentation.

This project also reinforced the importance of clean data, proper collaboration, version control, and modular task execution in real-world team settings. Every team member played a unique and valuable role in bringing the entire system to life.

Team Members & Contributions

- **VarunRaj** – Dataset Collection, Data cleaning, Verification, Documentation.
- **Anand Yadav** – Data Cleaning, Feature Engineering
- **Krish Gupta** – EDA & Visualization (Python)

- **Smriti Pandey** – Power BI Dashboard & Report
- **Shraddha Tiwari** – ML Model & Streamlit App
- **Roshni Sharma** – Alternate Regression Model & Colab Notebook