SAS Capstone Project

Objective: This project aims to answer business queries and find out descriptive statistics/insights about retail store orders basis on transactional data.

Data Description:

The description of the data is as mentioned below.
There are 2 files

- Orders.csv – This file contains the transaction level data of sales of a particular year where
    1. Order_id – order id of the sales and is unique per country
    2. Order_date – date on which order was placed
    3. Delivery_date – date on which order was delivered
    4. order_type – Source of Order
        - 1 – Retail, 2-Phone, 3-Internet, 99 – Invalid
    5. Product_id – ID of the product purchased
    6. Product category – Category of Product
    7. Quantity – Quantity of purchase
    8. Retail Price – Price of the
    9. Cost Price –
    10. Customer Country – Country of customer
    11. Customer Continent – Continent of Customer
    12. Customer_dob – DOB of the customer

- Country excel – This excel contains the country, their geographical information and population in the area
    1. "Countries"
        - Country_key – country code
        - Lat – Latitude
        - Lon – Longitude
        - Country Name – Name of the country
    2. Population
        - Country Name
        - Region
        - Population

Business Problem

- **Orders Frequency Analysis**: Find out and put your analysis in excel sheet
    - Which months have the highest and lowest total number of orders?
    - How many orders are distributed by each continent?
    - Within each continent how many orders were placed via retail, internet or phone?
- **Ship Days Summary**: Find out and put your analysis in excel sheet
    - How Many days on average does it take for an order to be delivered?
    - Are there any countries where shipment takes longer?
- **Profit Analysis by Customer Age:** Find out and put your analysis in excel
    - Which customer age group produces the highest median profit per order

Lab Environment: You need to have SAS EG or SAS University Edition installed on your machine.

Domain: Retail

Hints:

- Validate country lookup tables and put anomalies in a pdf
- Validate orders tables and put anomalies in a pdf
    - Categorical data analysis
        - Delivery date should be after order date
        - order_type valid values are 1,2 or 3
        - customer_country should be 2 upper case letters
        - customer_continents should be one of five continents
    - Continuous data analysis
        - Cost price, retail price, quantity
- Fix the anomalies found after data validation above by preparing the data and put the final data in an excel or permanent SAS dataset that will be used for analysis

---