Along with this document, you will receive a dataset with 3 tables:

- calendar
- customer_flight_activity
- customer_loyalty_history

This dataset describes an airline's customer activity. The table column description as follows:

Table	Field	Description
Customer Flight Activity	Loyalty Number	Customer's unique loyalty number
Customer Flight Activity	Year	Year of the period
Customer Flight Activity	Month	Month of the period
Customer Flight Activity	Total Flights	Sum of Flights Booked (all tickets purchased in the period)
Customer Flight Activity	Distance	Flight distance traveled in the period (km)
Customer Flight Activity	Points Accumulated	Loyalty points accumulated in the period
Customer Flight Activity	Points Redeemed	Loyalty points redeemed in the period
Customer Flight Activity	Dollar Cost Points Redeemed	Dollar equivalent for points redeemed in the period in CDN
Customer Loyalty History	Loyalty Number	Customer's unique loyalty number
Customer Loyalty History	Country	Country of residence
Customer Loyalty History	Province	Province of residence
Customer Loyalty History	City	City of residence
Customer Loyalty History	Postal Code	Postal code of residence
Customer Loyalty History	Gender	Gender

Customer Loyalty History	Education	Highest education level (High school or lower > College > Bachelor > Master > Doctor)
Customer Loyalty History	Salary	Annual income
Customer Loyalty History	Marital Status	Marital status (Single, Married, Divorced)
Customer Loyalty History	Loyalty Card	Loyalty card status (Star > Nova > Aurora)
Customer Loyalty History	CLV	Customer lifetime value - total invoice value for all flights ever booked by member
Customer Loyalty History	Enrollment Type	Enrollment type (Standard / 2018 Promotion)
Customer Loyalty History	Enrollment Year	Year Member enrolled in membership program
Customer Loyalty History	Enrollment Month	Month Member enrolled in membership program
Customer Loyalty History	Cancellation Year	Year Member cancelled their membership
Customer Loyalty History	Cancellation Month	Month Member cancelled their membership

Your task:

1. Using PySpark, extract the dataset into a dataframe, clean, and transform the dataset.
2. Using PySpark SQL, analyze the dataset and provide a minimum of 3 graphs.
3. Load your analysis result into a file (csv/parquet/json).
4. Please give documentation (.txt file) about your script, how to run, and your dataset analysis.

What to submit:

- Your python file (.py) that contains your answer.
- A folder that contains the exported graph and the data.

- A txt file that contains the documentation of your answer, and your analysis explanation.