**Challenge Name**: Customer Analysis

**Motivation:** The objective of this is to get the insights from the data. Datasets are available in the data directory. And the schema is as following:-

* Product (product\_id, product\_name, product\_type, product\_version, product\_price)
* Customer (customer\_id, customer\_first\_name, customer\_last\_name, phone\_number )
* Sales (transaction\_id, customer\_id, product\_id, timestamp, total\_amount, total\_quantity )
* Refund (refund\_id, original\_transaction\_id, customer\_id, product\_id, timestamp, refund\_amount, refund\_quantity)

**Deliverables**:

* Write down the data quality issues with the datasets provided and steps performed to clean (if any).
* Display the distribution of sales by product name and product type.
* Calculate the total amount of all transactions that happened in year 2013 and have not been refunded as of today.
* Display the customer name who made the second most purchases in the month of May 2013. Refunds should be excluded.
* Find a product that has not been sold at least once (if any).
* Calculate the total number of users who purchased the same product consecutively at least 2 times on a given day.

**Non-Functional features:** Scalability, Usability**,** Availability

**DataSets**: Attached.

