# SQL

Given a database with 3 tables as follows:

**category** (index = **id**)

|  |  |
| --- | --- |
| **id** | **name** |
| 1 | Makanan Instan |
| 2 | Fresh |
| 3 | Makanan Ringan |
| 4 | Rokok |
| ... | ... |
| 99 | Minuman Kemasan |

**product** (index = **id**)

|  |  |  |
| --- | --- | --- |
| **id** | **name** | **category\_id** |
| 1 | Filma Minyak Goreng | 7 |
| 2 | Beras si Geulis | 8 |
| 3 | Kapal Api SpecialMix | 15 |
| 4 | Ladaku MericaBubuk | 10 |
| ... | ... |  |
| 99 | Djarum Gold | 4 |

**sales\_order** (index = **trx\_id**)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **trx\_id** | **trx\_line\_id** | **product\_id** | **trx\_state** | **order\_value** | **created\_at** |
| 1 | 1 | 3 | Delivered | 100,000 | 2020-10-05 21:57:26 |
| 2 | 1 | Cancelled | 300,000 | 2020-08-15 15:05:55 |
| 3 | 1 | Cancelled | 1,500,000 | 2020-09-18 20:02:58 |
| 2 | 4 | 5 | Delivered | 50,750,000 | 2020-11-05 13:40:17 |
| 5 | 1 | Cancelled | 7,750,000 | 2020-10-24 12:49:40 |
| 3 | 6 | 4 | Delivered | 5,500,000 | 2020-09-21 9:42:39 |
|  | ... | ... | ... | ... | 2020-09-25 8:16:32 |
| 99999 | 99999 | 7284 | Delivered | 25,000,000 | 2020-09-14 20:46:29 |

1. Write SQL queries to:
   * Get top 3 products with the highest total sales value in October 2020 in each category. Show the category name, product name, and total sales value.

Total Sales Value = Order Value from delivered transaction (trx\_state = ‘Delivered’).

* + Get all products with sales value lower than average sales value of products within the corresponding category. Show category name, product name, product sales value, and average sales value in corresponding category.
  + Calculate sales lost and sales lost rate of each category, sort by sales lost descending. Show category name, sales lost, and sales lost rate.

Sales lost = Order Value from Cancelled Transactions.

Sales Lost Rate = (Order value from cancelled transactions) / (Total order value from all transactions

Graphical user interface, application

Description automatically generated

# MODELLING

# Given data as follows:

|  |  |  |
| --- | --- | --- |
| application\_record.csv |  |  |
| Feature name | Explanation | Remarks |
| ID | Client number |  |
| CODE\_GENDER | Gender |  |
| FLAG\_OWN\_CAR | Is there a car |  |
| FLAG\_OWN\_REALTY | Is there a property |  |
| CNT\_CHILDREN | Number of children |  |
| AMT\_INCOME\_TOTAL | Annual income |  |
| NAME\_INCOME\_TYPE | Income category |  |
| NAME\_EDUCATION\_TYPE | Education level |  |
| NAME\_FAMILY\_STATUS | Marital status |  |
| NAME\_HOUSING\_TYPE | Way of living |  |
| DAYS\_BIRTH | Birthday | Count backwards from current day (0), -1 means yesterday |
| DAYS\_EMPLOYED | Start date of employment | Count backwards from current day (0). If positive, it means the person currently unemployed. |
| FLAG\_MOBIL | Is there a mobile phone |  |
| FLAG\_WORK\_PHONE | Is there a work phone |  |
| FLAG\_PHONE | Is there a phone |  |
| FLAG\_EMAIL | Is there an email |  |
| OCCUPATION\_TYPE | Occupation |  |
| CNT\_FAM\_MEMBERS | Family size |  |

# 

# Graphical user interface, application Description automatically generated

|  |  |  |
| --- | --- | --- |
| credit\_record.csv |  |  |
| Feature name | Explanation | Remarks |
| ID | Client number |  |
| MONTHS\_BALANCE | Record month | The month of the extracted data is the starting point, backwards, 0 is the current month, -1 is the previous month, and so on |
| STATUS | Status | 0: 1-29 days past due 1: 30-59 days past due 2: 60-89 days overdue 3: 90-119 days overdue 4: 120-149 days overdue 5: Overdue or bad debts, write-offs for more than 150 days C: paid off that month X: No loan for the month |

# Build a machine learning model to predict if an applicant is 'good' or 'bad' client. The definition of 'good' or 'bad' is not given. You should use some techniques to construct the label.

# please explore the data and give some insight from your exploration. Then, story telling your model result.

# Given data as follows:

|  |  |  |
| --- | --- | --- |
| Sales\_Groceries.csv |  |  |
| InvoiceNo | Invoice number | a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation. |
| StockCode | Product (item) code | a 5-digit integral number uniquely assigned to each distinct product. |
| ProductName | Product (item) name. |  |
| Quantity |  | The quantities of each product (item) per transaction |
| InvoiceDate | Invoice Date and time | Numeric, the day and time when each transaction was generated. |
| UnitPrice | Unit price | Product price per unit in poundsterling. |
| CustomerID | Customer number | a 5-digit integral number uniquely assigned to each customer |
| Country | Country name | the name of the country where each customer resides. |

# Build a machine learning model for predicting daily revenues from an online retail.

# what do you think is the most important step(s) during forecasting modeling?

# You are expected to present the code and the explanation later during technical interviews.

# Goodluck!