customer-analysis-5

Share

300M Retail Data Analysis Project Documentation

Created by Li Guangyu
Last updated: just a moment ago by Helena • 3 min read

Project Description:

This project aims to use analytic tools including Excel, Python, Tableau, etc to analyze the potential regular pattern in the provided retail sales data and give suggestions to the operation for the incoming sale performance. The expected items include but are not limited to the populator products, the correlation between order amount and order time, and the department's order frequency.

Data analysis mainly includes the following stages: identify the issues, data collection, cleaning data, data modeling, and data visualization. This project of the retail sales data mainly focuses on data transformation and ETL, data modeling, and project presentation three parts supported by machine learning, Python, Tableau, etc. We can discover useful information in the analysis of data and make the regular pattern support company's operation.

This project would also involve many methodologies like database fundamentals, data analysis basics, python basics, industry framework.

Content

This project contains several important steps: data analysis basics, Machine learning, project presentation, data transformation, and ETL, applied modeling process, and a recommendation system.

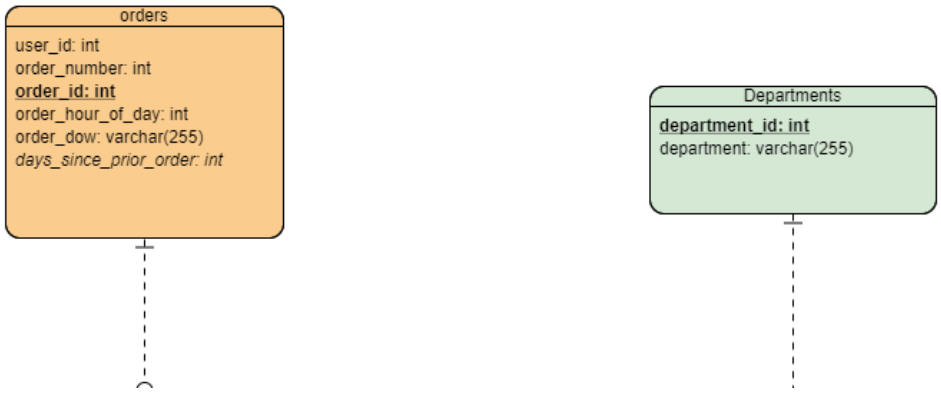
Project goal:

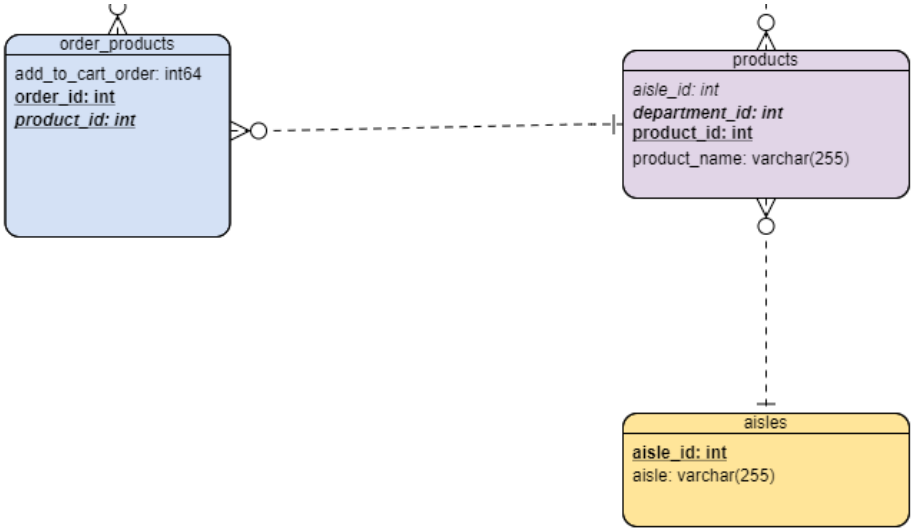
This project aims to use data mining and machine learning technics find the potential regular pattern in customer behaviours and help companies to better understand clients' behaviours and support the decision-making in operation and targeted build a product recommendation engine to increase sales performance and meet customers' demand.

Implementation:

- The use of data storage, putting data to AWS, and define a schema for raw data
- The use of SQL for basic data analysis and build data relationships
- The use of Python (Numpy, Pandas, Scikit Learn, matplotlib, seaborn) for data analysis and visulization
- The use of Tableau or Power BI to build a dashboard and present the relationship in different items
- Unsupervised learning for customer segmentation and visualization
- The discover of hidden patterns or data groupings without the need for human intervention
- Build a machine learning model for predictions
- The final outcome would be the recommendation system.

ER model:





Data dictionary

Aisles					
Field Name	**Data Type**	**Data Format**	**Field Size**	**Description**	**Example**
aisle_id	int64	NNNNN	64	Unique number ID for all products	8
aisle	varchar(255)	NNNNN	255	Unique type in products	Soap

Departments					
Field Name	**Data Type**	**Data Format**	**Field Size**	**Description**	**Example**
department_id	int64	NNNNN	64	Unique number ID for all departments	5
department	varchar(255)	NNNNN	255	Unique name for all departments	fruit

Order_products					
Field Name	**Data Type**	**Data Format**	**Field Size**	**Description**	**Example**
order_id	int64	NNNNN	64	Unique number ID for all orders	2539329
product_id	int64	NNNNN	64	Unique number ID for all products	6
add_to_cart_order	int64	NNNNN	64	the order that put in the cart	3

Orders					
Field Name	**Data Type**	**Data Format**	**Field Size**	**Description**	**Example**
order_id	int64	NNNNN	64	Unique number ID for all orders	2539329
user_id	int64	NNNNN	64	Unique number ID for all users	23
order_number	int64	NNNNN	64	Sequential number for all orders in one user_id	54
order_dow	int64	NNNNN	64	the order day of week, weekday	3 (0-6)
order_hour_of_day	int64	NNNNN	64	the order hour in a day	12 (0-24)
days_since_prior_order	float64	NNNNN	64	the days since the order date last time	29 (0-31)

Products					
Field Name	**Data Type**	**Data Format**	**Field Size**	**Description**	**Example**
product_id	int64	NNNNN	64	Unique number ID for all products	6
product_name	varchar(255)	NNNNN	255	Product name	nectar
aisle_id	int64	NNNNN	64	Unique number ID for all products	8
department_id	int64	NNNNN	64	Unique number ID for all departments	5