CSC3170 project

**数据集：**

<https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>

**时间线：**

5.8 8:00 p.m. ppt上交ddl

5.9 final pre

5.11 11:00 p.m. final report ddl

**细化任务：**

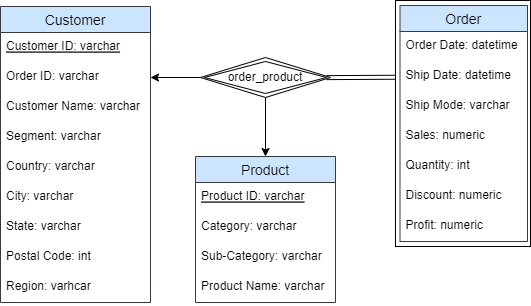
1. 前端网页（用于输入查询所需的数据库）（非必要但挺多组都做了）负责：wzh
2. 分析可能的需求（例如统计商品总数 库存数量 商品一个时间段的销售量之类的 我觉得可以去专门看看相关的分析 这样我们后面SQL和数据库的设计会更合理一些？）负责：zyf zyb
3. 分析数据的entity以及他们的关系，产出E-R图：swl gh
4. “Convert the E-R diagrams to relational schemas”， 同时指出PK, FK, functional dependencies，multivalued dependencies
5. 生成实例SQL语句（“Suggest which data fields of the relational schemas should be indexed or hashed, and explain your decision”）
6. REPORT相关：

* Introduction and motivation
* Design and implementation
* Conclusion and self-evaluation
* References
* Appendices

**Analysis：**

**最基础的统计问题：**

* 哪个顾客有最高的消费次数/最高的消费额？
* 哪个商品卖的最好？（或者TOP10）
* 最高消费额的TOP10city
* 最高盈利额的TOP10city
* 哪个品类（category）盈利最多/购买次数最多



Relational schema:

Customer {CustomerID, Order ID, Customer\_Name, Segment, Country, City, State, Postal\_Code, Region}

Product {ProductID, Category, Sub\_Category, Product\_Name}

Order {Order\_Date, Ship\_Date, Ship\_Mode, Sales, Quantity, Discount, Profit}