**Part\_A:資料初步統計**

***Stage1:分析會員的消費維度（按時間變化）***

**基礎維度：**

A，累計消費天數、累計消費單數、累計消費件數、累計消費金額、消費品類數目、消費品牌數目、最後消費時間距離、第一次消費間隔天數、

B，平均消費間隔天數

C，積分

**衍生維度:**

品類的單數、額度

**分析分佈及關係**：

對數正態的有：

平均消費間隔天數、累計消費金額、平均每件消費金額、平均每天消費金額；

較強正相關關係有：

累計消費件數-累計消費金額、平均每件消費金額-平均每天消費金額；

***Stage2:關聯會員品牌品類資料（按時間變化）***

品牌偏好（單量，金額）

品牌單數,額度,平均品牌單數衡量會員的品牌集中度,品牌大類的單數分佈衡量

***Stage3:關聯會員的個人資料（不按時間變化）***

性別、擁有車情況、年齡、戶籍地區

**分析分佈及關係**：

對數正態的有：年齡

***Stage4:計算衍生維度***

平均每天消費金額、平均每單消費金額、平均每件消費金額、天均單、單均件

***Stage5:按照維度劃分標籤***

分類型變數：是否有車、性別、地區

**區間分佈及主要劃分：**

偏分類型（非正態分佈，主要數目集中在少數值上，分別以最小值、占比在50%、75%劃分區間）：

消費天數為1 的會員、總天數（1,2~4,5~11,12以上）、總單數（1,2~6,7~15,16以上）、總件數（1,2~7,8~18,19以上）、品類數（1,2~3,4~5,6以上）、品牌數（1,2~5,6~11,12以上）、天均單（1, ~1.5,大於1.5）、單均件（1， ~1.222，大於1.222）

偏數值型（正態分佈）：

消費大於一天的平均間隔天數、五萬以內的銷售總額、0~60的年齡(像年轻偏度)、一萬以內的天均額、一萬以內的單均額、一萬以內的件均額、积分、剩余积分

均匀分布（随机性强，通过变量关联划分）

首次消費距離天數、後消費間隔天數、

分布不纯（正态部分区间差异大）

返利积分

***Stage6：各類別會員數量***

**Part\_B:數據主要分析類型**

***Stage7:分析維度差異及選取分析變數）***

**分*析維度差*異：**

A，分类之间（卡方检验）：

线性相关：天数-（单量，件数，品类数，品牌数），单量-（件数，品类数，品牌数）

件数-（品类数，品牌数），品类数-（品牌数）；

方差递减:（天数，单量，件数，品类数，品牌数）增加-（天均单，单均件）趋于中值稳定，均值为1的越少；

B，连续变量之间（相关分析）：

剔除件均额，与单均额极高相似（相关系数为1）

C,分类与连续变量之间：

差异明显的组合：（天数，单数，品类数，品牌数）-（平均时间间隔，总额度，积分，剩余积分，天均额，单均额，件均额，（返利积分））

差异不明显的组合：

（天数，单数，品类数，品牌数）-年龄，（返利积分）

结果：选取天数多的天均额、单均额、件均额作为各自的平均值参考点，品类数、品牌数为1是收到个别价格区间的影响较大，

D，其他：

天数-品牌数，偏重品牌的多元化；

***Stage7\_sub：(划分随机性强的变量）***

均匀分布（随机性强）：

1. 首次消費距離天數（以871划分）：

与天数（单量，件数，品类数，品牌数）,消费天数多的主要集中在首次消费间隔天数

871天上；

与（天均单，单均件）, （天均单，单均件）少的在首次消费间隔天数少的比重大，首次消费间隔天数多的（天均单，单均件）主要集中在2class；

1. 最後消費間隔天數（收到活动的影响大，保留原数据）：

与天数（单量，件数，品类数，品牌数）,消费天数多的主要集中在最后消费间隔天数

127天以下；

***Stage8:會員資料的回歸分類分析***

**重複返店會員的相關因素:**

維度同stage1 的消費維度

分類標籤：

店慶因素不敏感會員。

**購買某品牌的相關因素**

**品牌的關聯銷售**

***Stage6:會員資料的其他分類分析***

根據消費記錄計算相似消費客戶

根據消費記錄計算相似品牌