作成日：2024年4月21日

**2023年度データアナリティクスII**

**莊司担当モジュール**

**課題レポート**

提出期限：2024年5月20日

所属：静岡大学情報学部 XXXX学科

学籍番号：xxxxxx

氏名：xxxxxx

## 課題概要

　関係データベースを利用して購買データを管理している仮想の小売店（杏森堂）から，データ分析の依頼を受けたとする．授業で使用したDeepnote内のデータと分析環境を利用して，下記課題に回答しなさい

## 課題0（配点15%）

初回授業で配布されたZipファイルには，売上に関する以下の4つのファイルが格納されている：

* 購買記録に関するデータ（2019年1月〜2019年7月の売上履歴）
* 顧客台帳データ（手入力で店舗が管理している顧客台帳）

2つのファイルを用いて，以下に関する表を作成し、スクリーンショットを貼れ。

* 1. **縦方向にある年のある月，横方向に商品を並べた，商品ごとの年月別売上情報**

（ここにスクリーンショットを貼る）

* 1. **縦方向にある年のある月，横方向に地域（市）を並べた，各市ごとの年月別売上情報**

（ここにスクリーンショットを貼る）

* 1. **集計期間中（2019年1月〜2019年7月）に杏森堂で購買行動をしていない顧客のリスト**

（ここにスクリーンショットを貼る）

## 課題1（配点25%）

　データベースを整備した杏森堂から，再度データ分析の依頼を受けたとする． 購買履歴が格納されたデータベースを用いて，これまでに何名の顧客が何回（何日）ショッピングを行ったのかを分析したい．

　レシート明細テーブル（receipt）を用いて顧客の購買頻度を分析し，以下の項目について分析結果を表示するSQL文を書き，そのSQL文の意味を解説し，実行結果を示しなさい：

* 購買頻度（これまでに店舗を利用した日数）
* 購買頻度に対応する顧客の数
* 該当する購買頻度以下の顧客数の累積値

ただし，顧客ID（customer\_id）が"Z"から始まるのものは非会員を表すため，除外して分析すること．

## 課題1に対する回答

### 実行したSQL

### 実行したSQL文の解説

### 実際の実行結果（スクリーンショットでも、テーブル形式でもよい。）

## 課題2（配点25%）

　課題2-1，課題2-2のいずれかを選択し，結果を得るためのSQL文を書き，SQL文の解説と、実行結果を示しなさい．

課題2-1: 月別売上の対昨年比

　レシート明細テーブル（receipt）には2017年1月から2019年10月の間の購買情報が記録されている．2017年から2019年までの期間の売上を把握するために，1ヶ月ごとに以下の情報を集約表示せよ：

* 年月（year\_month）
* 購買回数（purchase\_freq）
* 購買1回あたりの平均購入額（avg\_amount\_per\_purchase）
* 月間売上高（total\_amount）
* 当該月の前年売上高（total\_amount\_last\_year）
* 売上の対前年比（ratio）

課題2-2: ABC分析

　ABC分析は販売戦略を考えるために，売上によって商品をランク付けする手法である． 一般に，ABC分析では売上総額の

* 上位0〜70％の商品をAランク
* 上位70〜90％の商品をBランク
* 上位90〜100%の商品をCランク

とするランク付けを行う．

　商品カテゴリ「菓子（category\_major\_cd = 08）」に属する商品について，小区分（category\_small\_name）ごとに売上を集計し，菓子カテゴリの売上総額に占める割合（構成比）を計算せよ． また，売上がN位の小区分の行には売上額上位N位までの構成比累計，および構成比累計に基づくABC分析のランク付け結果も表示せよ．

## 課題2に対する回答

### 実行したSQL

### 実行したSQL文の解説

### 実際の実行結果（スクリーンショットでも、テーブル形式でもよい。）

## 課題3（配点25%）

　課題3-1から3-4の4題から1題を選択し回答しなさい．なお，課題3については，2題以上解いた学生には1題につき最大10％を加点する．

課題3-1

　顧客ごとに購買頻度（購買回数）を求め，購買頻度の分布を説明するための任意の形式のグラフを作成せよ．また，グラフから読み取れることを文章で記述せよ．また，なぜその形式のグラフでこのデータを説明すべきなのか，理由を説明せよ．ただし，顧客ID（`customer\_id`）が"Z"から始まるのものは非会員を表すため，除外して分析すること．

　なおレポートには，顧客ごとに購買頻度（購買回数）を求めるためのSQL文と解説も付しておくこと（SQL文の実行結果は掲載する必要なし）．

課題3-2

　都道府県別に年月別の売上総額を求め，その売上総額の変化を比較するための任意の形式のグラフを作成せよ．また，グラフから読み取れることを文章で記述せよ．また，なぜその形式のグラフでこのデータを説明すべきなのか，理由を説明せよ．

　なおレポートには，都道府県別の年月別売上総額を求めるためのSQL文と解説も付しておくこと（SQL文の実行結果は掲載する必要なし）．

課題3-3

　顧客一人あたりの売上総額の分布を，千葉県の店舗ごとに比較するための任意の形式のグラフを作成せよ．また，グラフから読み取れることを文章で記述せよ．また，なぜその形式のグラフでこのデータを説明すべきなのか，理由を説明せよ．ただし，顧客ID（`customer\_id`）が"Z"から始まるのものは非会員を表すため，除外して分析すること．

　なおレポートには，店舗ごとにその店舗を利用した各顧客の売上総額を集計するSQL文と解説も付しておくこと（SQL文の実行結果は掲載する必要なし）．

課題3-4

　千葉県にある店舗間で顧客が購入する菓子の傾向が異なるかどうかを分析したい．千葉県にある店舗と菓子の中カテゴリ（cateogory\_medium\_cd）ごとに菓子の購入量（個数）を集計し，購入された菓子の中カテゴリの内訳を店舗間で比較するためのグラフを作成せよ．グラフの形式は任意の形式でよい．また，グラフから読み取れることを文章で記述せよ．また，なぜその形式のグラフでこのデータを説明すべきなのか，理由を説明せよ．

　なおレポートには，購入量を集計するためのSQL文も付しておくこと（SQL文の実行結果は掲載する必要なし）．

## 課題3に対する回答

xxxxx

## 課題4（配点10%）

　データアナリティクスIIの莊司担当パートについて，良かった点および改善すべき点を記しなさい．

## 課題4に対する回答

良かった点

xxxx

改善すべき点

xxxx