Кадарметов Дмитрий

Группа: БИВТ-ВП-23

**Лабораторная работа 1.**

**Тема.**Сбор и подготовка данных.

**Цель**: освоить навык предварительной подготовки данных.

**Этапы выполнения работы:**

1. Определить и описать кратко предметную область. Сформулировать гипотезу или цель анализа, которые можно в дальнейшем проверить на основе анализа выбранных данных.
2. Для выбранной предметной области зафиксировать анализируемый объект и его характеристики (2 или 3). Найти соответствующие данные для анализа. Указать источник данных.
3. Описать генеральную совокупность. Определить и описать способ построения выборки из генеральной совокупности объекта исследования.
4. Представить данные выборки в табличном виде: строки - экземпляры объекта, столбцы - данные характеристик.
5. Выполнить и зафиксировать в отчете результат необходимых операций предобработки данных (в случае отсутствия необходимости выполнения конкретной операции, указать - не требуется):

         -     разделение данных,

         -     приведение данных к одинаковым единицам измерения,

         -     преобразование к унифицированной лексике,

         -     объединение данных из разных источников,

         -     соединение данных из разных источников,

         -     агрегация данных,

         -     заполнение отсутствующих числовых значений,

         -     очистка данных.

1. В соответствии с имеющимися данными построить график для каждого набора данных (для каждой характеристики объекта).
2. Построить диаграмму рассеяния (точечная диаграмма) для набора данных.
3. Сделать вывод.
4. Подготовить и загрузить отчет.

# Отчет о проделанной работе

* 1. Предметная область – продажи кофейных напитков в кофейнях. Она включает в себя процессы, связанные с приготовлением и реализацией кофе, чая и других напитков, а также продажей сопутствующих товаров (выпечка, десерты) в условиях кофейни.

Цель работы – получить базовое представление о динамике продаж кофейных напитков в течение года чтобы определить наиболее благоприятное время для открытия кофейни.

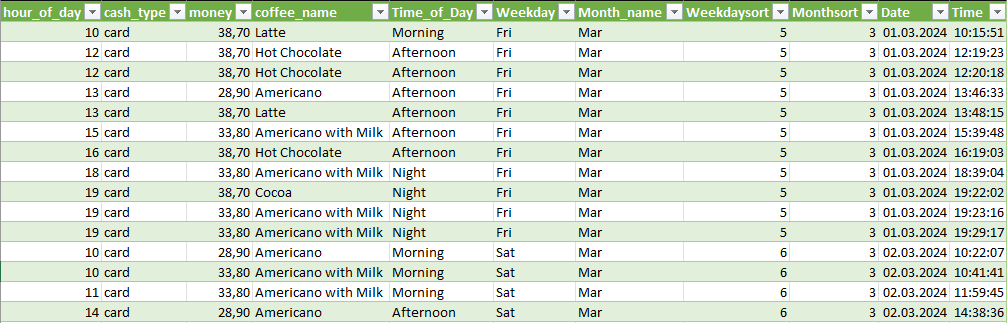
* 1. В рамках настоящей работы мы будем анализировать только продажи кофейных напитков, поскольку они являются основным товаром, приносящей предприятию прибыль.

Кофейные напитки бывают разных видов (американо, капучино, и др.), могут приобретаться в разное время суток (утро, день, вечер, ночь), в разное время года (осень, зима, весна, лето), с помощью разных типов оплаты (наличные, карта), за разную стоимость.

Источник данных для анализа – сайт Kaggle.

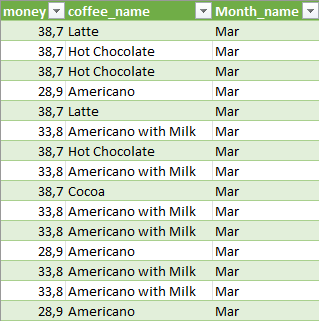
Ссылка на набор данных: https://www.kaggle.com/datasets/navjotkaushal/coffee-sales-dataset/data

* 1. Генеральная совокупность – все кофейные напитки, проданные за все время в кофейнях. Вид выборки: простая случайная - статистический метод, при котором каждое наблюдение в большой совокупности имеет равные шансы попасть в выборку.
  2. Данные выборки:

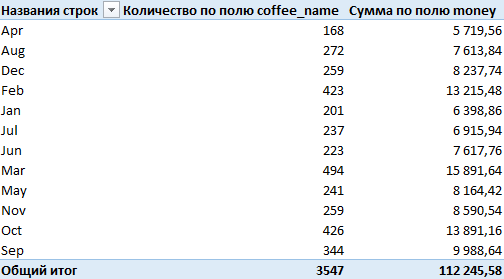


* 1. Предобработка данных:
* разделение данных – не требуется
* приведение данных к одинаковым единицам измерения – не требуется
* преобразование к унифицированной лексике – не требуется
* объединение данных из разных источников – не требуется
* соединение данных из разных источников – не требуется
* агрегация данных – не требуется
* заполнение отсутствующих числовых значений – не требуется
* очистка данных – удаление столбцов, дублирующих информацию, и столбцов с данными, не относящимися к цели анализа, а именно: час дня, тип оплаты, время дня (утро, день, вечер), день недели, номер дня недели, номер месяца, точная дата и точное время совершения покупки.

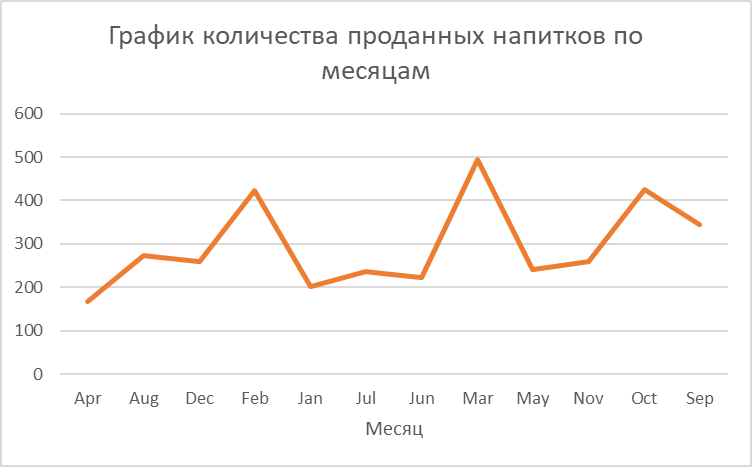
Таблица после предобработки:

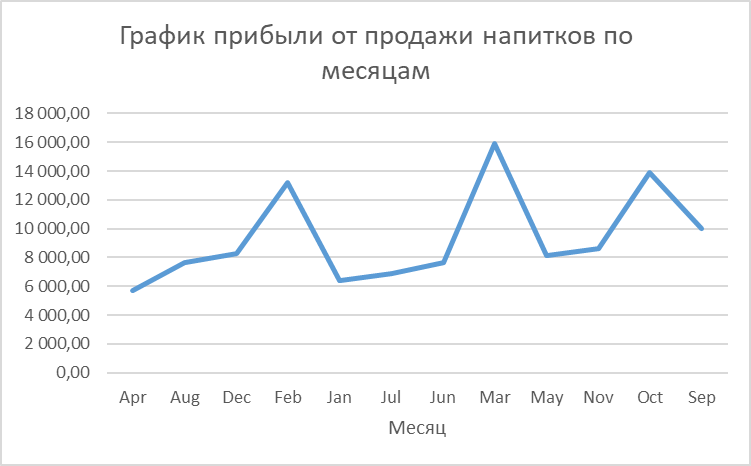


После предобработки на основе полученной таблицы была построена сводная таблица:

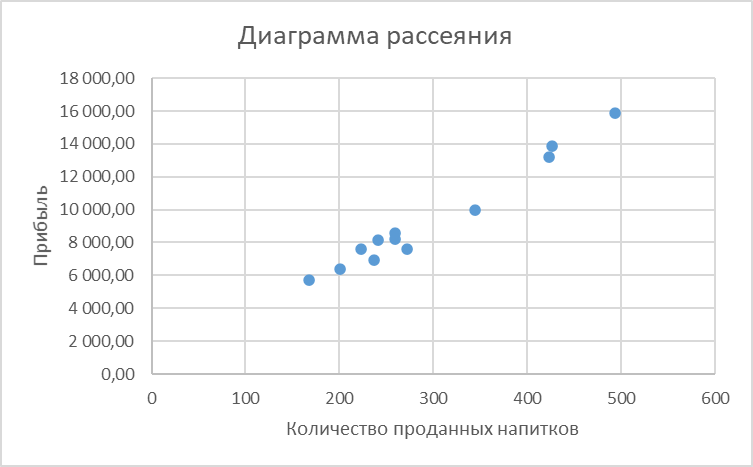


* 1. На основе построенной сводной таблицы были построены графики:





* 1. Диаграмма рассеяния для набора данных



* 1. Вывод:

В ходе работы были проведены сбор и предварительная подготовка данных для анализа, после чего по обработанным данным были построены графики и диаграмма рассеяния.

По графикам и диаграмме можно заметить что:

1. Пик продаж кофейных напитков – это месяцы: февраль, март, октябрь и сентябрь, причем в марте количество продаж особенно отличается от остальных месяцев, что можно видеть на диаграмме рассеяния.
2. Графики количества проданных напитков по месяцам и прибыли от продаж по месяцам практически совпадают по своей форме, из чего можно сделать вывод, что соотношение между количеством приобретаемых напитков по виду напитка остается достаточно постоянным в течение года. Небольшие отличия в форме графика наблюдаются в июле и августе.