# Отчет по лабораторной работе 1

# Тема: Сбор и подготовка данных

# Цель: освоить навык предварительной подготовки данных

# Выполнил: Кадарметов Дмитрий Николаевич

# Группа: БИВТ-ВП-23

# Этап 1

Предметная область – продажи кофейных напитков в кофейнях. 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.

Для анализа было взято два готовых набора данных о продажах кофейных напитков.

Из-за особенности полученных данных не представляется возможным анализировать типы оплаты (поскольку первый набор данных сформирован только на основе покупок, совершенных по карте) и стоимость покупки (поскольку в первом наборе данных стоимость приведена в местной валюте, но регион покупки не указан). В следствие этого предполагается анализировать вид кофейного напитка, время суток и время года, в которые была совершена покупка.

# Этап 2

Предметная область включает в себя процессы, связанные с приготовлением и реализацией кофе, чая и других напитков, а также продажей сопутствующих товаров (выпечка, десерты) в условиях кофейни

Гипотезы для проверки:

1. Самые популярные виды кофе: капучино, латте, американо
2. Пик продаж приходится на середину дня
3. В холодное время года частота продаж увеличивается.

Независимые переменная – количество операций. Зависимые переменные: вид кофейного напитка, время суток, время года.

# Этап 3

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.

# Этап 4

Примеры необработанных таблиц: см. в дополнительных файлах к отчету

# Этап 5

После обработки и объединения, таблицы приняли следующий вид:

|  |  |  |
| --- | --- | --- |
| hour\_of\_day | coffee\_name | Date |
| 10 | Latte | 01.03.2024 |
| 12 | Hot Chocolate | 01.03.2024 |

Для этого было произведено удаление столбцов таблицы, данные в которых не представляют интереса для проверки гипотезы, а также изменение порядка столбцов для приведения таблиц к единообразию.

# Этап 6

Построенные диаграммы:

# Этап 7

Диаграммы рассеивания:

# Этап 8

Выводы:

1. Гипотеза подтвердилась, однако одним из самых популярных кофейных напитков было американо с молоком, которое не учитывалось в гипотезе.
2. Гипотеза оказалась ошибочна – пик продаж пригодится на утро (10 часов утра).
3. Гипотеза оказалась ошибочна – частота продаж наивысшая весной и осенью.

Возможные выводы по соответствующим гипотезам:

1. Ассортимент кофейни обязательно должен содержать латте, американо и капучино, поскольку это наиболее популярные виды кофейных напитков.
2. Утром и днем владельцам кофейни следует ожидать наибольшее число клиентов.
3. Для открытия новой кофейни следует выбирать или весну, или осень, поскольку пик продаж приходится именно на это время года.