

**Московский государственный технический  
университет им. Н.Э. Баумана**

Факультет «Информатика и системы управления»  
Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Технологии машинного обучения»

Отчет по лабораторной работе №1

«Разведочный анализ данных. Исследование и визуализация данных.»

Выполнил:  
студент группы ИУ5-62Б  
Перова Анна

Проверил:  
преподаватель каф. ИУ5  
Гапанюк Ю.Е.

Москва, 2022 г.

## Описание задания:

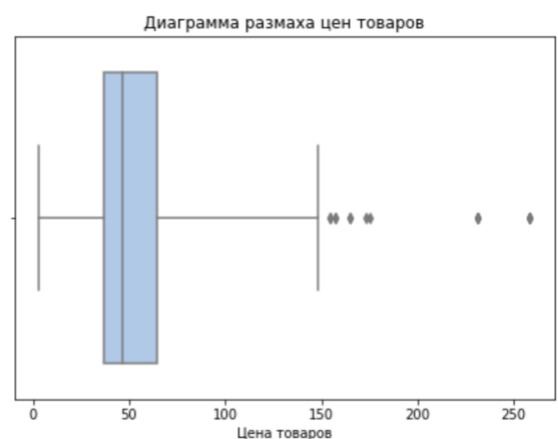
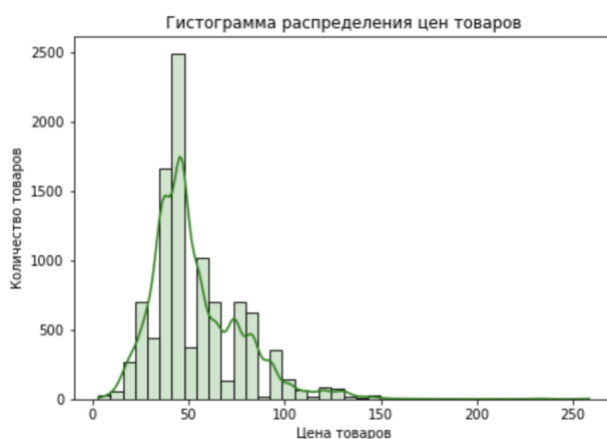
- Выбрать набор данных (датасет).
- Создать ноутбук, который содержит следующие разделы:
  1. Текстовое описание выбранного Вами набора данных.
  2. Основные характеристики датасета.
  3. Визуальное исследование датасета.
  4. Информация о корреляции признаков.

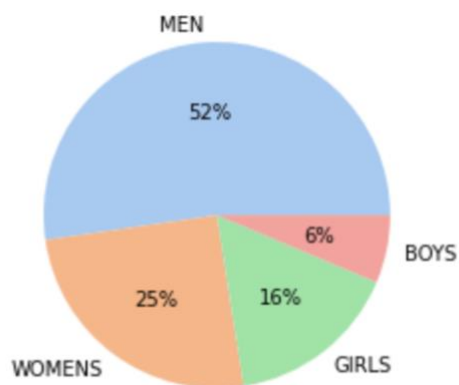
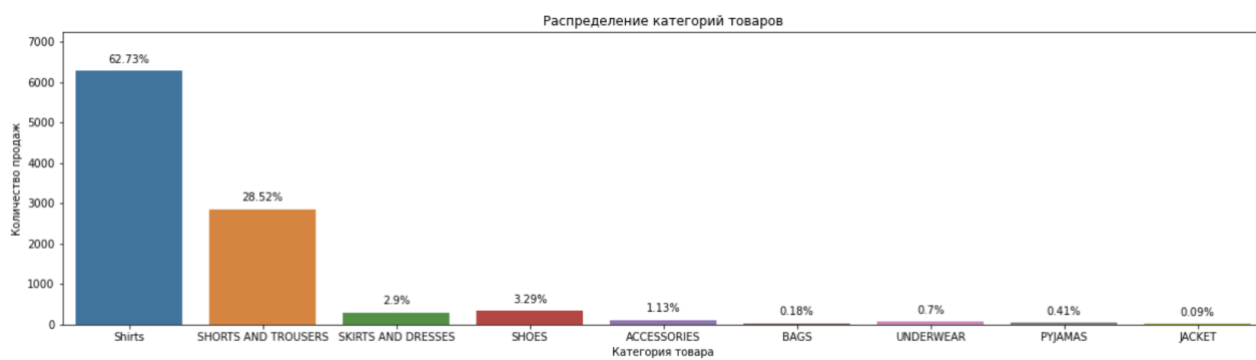
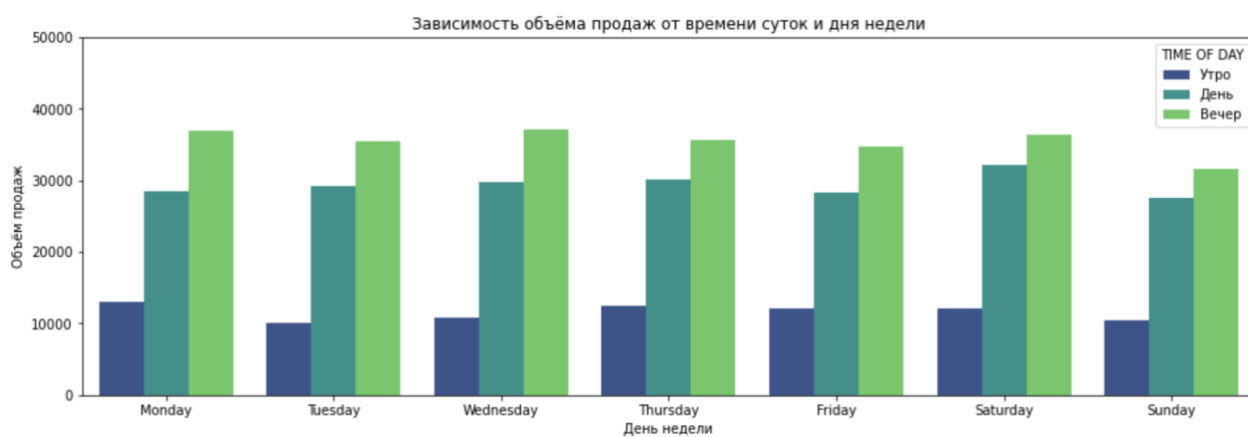
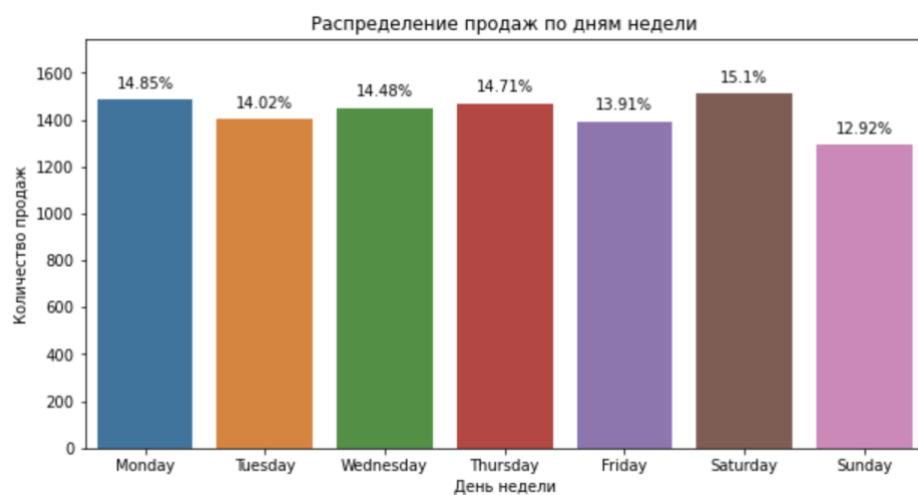
## Описание датасета

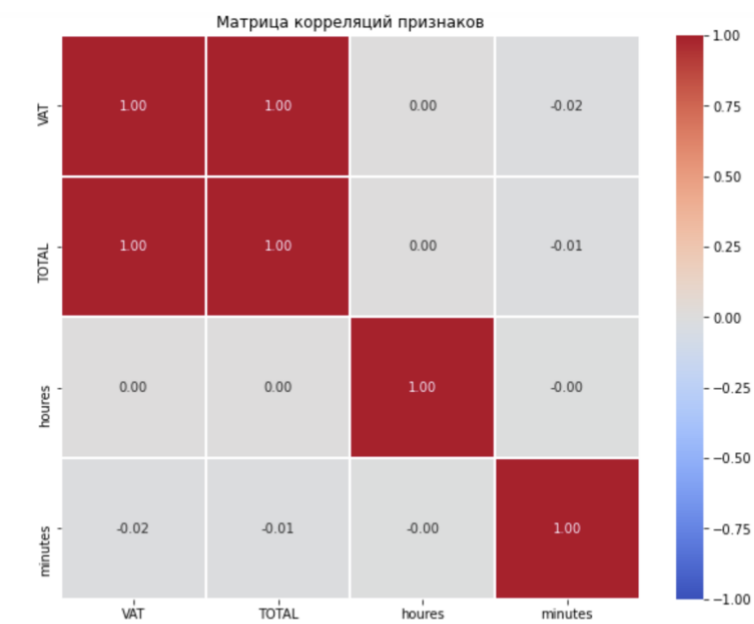
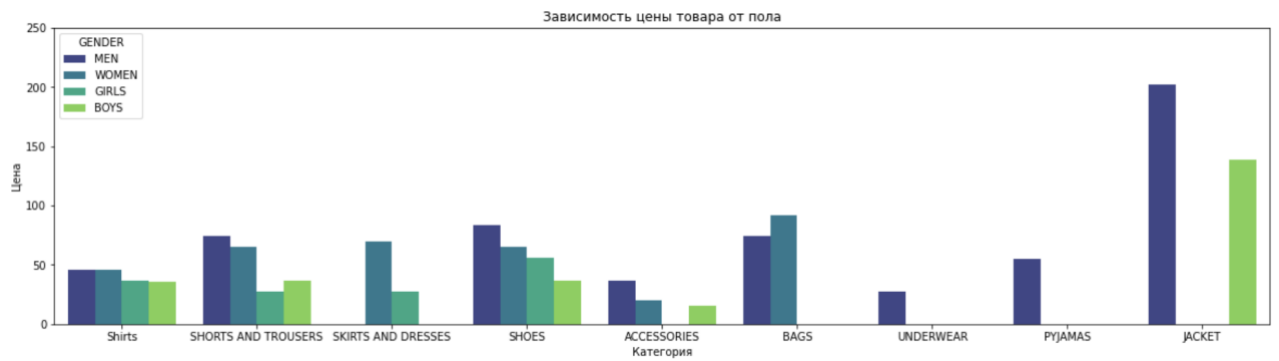
Датасет `SHOP-SALES.xlsx` содержит информацию о продажах магазина.

Параметры покупки:

- *Date* - дата покупки,
- *Hour* - час совершения покупки,
- *Product* - тип купленного товара,
- *Gender* - пол,
- *Color* - цвет товара,
- *Size* - размер товара,
- *Price* - цена товара без учета НДС,
- *Vat* - НДС,
- *Total* - Цена товара с НДС.







Видим линейную зависимость между значением НДС и итоговой стоимостью товара. Остальные признаки не коррелируют между собой.

Теперь найдём коэффициенты корреляции между признаками  $X$  и целевой переменной  $y$ , чтобы определить влияние признаков на стоимость ноутбуков и сравнить результаты с закономерностями, выявленными ранее при визуализации распределений признаков  $X$  и их связей с целевой переменной  $y$ .

