

**«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ИМЕНИ М.В. ЛОМОНОСОВА**

Факультет Вычислительной Математики и Кибернетики

Кафедра исследования операций

Отчет по заданию №4

Студенты:	Липатова Александра Смирнов Михаил Разумова Вера
-----------	--

Преподаватель:	Гусева Юлия
----------------	-------------

Группа:	312
---------	-----

Постановка задачи.

После оглушительного успеха в освобождении Астапора, Миэрина и Юнка от власти работоторговцев Дейенерис Бурерожденная открыла себе доступ к Летнему морю, а следовательно – путь в Вестерос.

Для ведения войны с Семью Королевствами нужно оружие, а для оружия нужна сталь. Нет никаких сомнений в кузнечном искусстве Безупречных, однако поставщики стали не столь надежны. Два основных поставщика стали – это Westeros Inc. и Harpy Co. На протяжении нескольких месяцев мы закупаем сталь у обеих компаний, и каждая из них предлагает ощутимую скидку при заключении эксклюзивного договора на поставку.

Советник королевы Тирион Ланнистер знает о твоём умении принимать взвешенные рациональные решения и просит помощи в объективном решении вопроса о том, с какой из компаний следует заключить эксклюзивный договор на поставку стали. У Тириона есть записи о производстве мечей каждым из кузнецов-безупречных, а также данные о количестве сломанных мечей в каждый из месяцев ведения боевых действий.

Входные данные из CSV таблицы, отсортированные по ID кузнеца:

- Supplier (поставщик)
- Produced (количество произведенных мечей в месячной поставке)
- Report date (месяц поставки)
- Defects (количество сломанных мечей в каждой поставке)
- Unsullien ID (ID кузнеца)

Необходимо провести разведывательный анализ данных с целью ответа на вопрос: "С каким из поставщиков стали следует заключить договор?". То есть задача сводится к транспортной задаче по поиску максимально выгодного поставщика.

ОСНОВНЫЕ ПОНЯТИЯ.

Разведочный анализ данных (англ. *exploratory data analysis*, EDA) — анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей, зачастую с использованием инструментов визуализации.

Понятие введено математиком Джоном Тьюки, который сформулировал цели такого анализа следующим образом:

- максимальное «проникновение» в данные
- выявление основных структур
- выбор наиболее важных переменных
- обнаружение отклонений и аномалий
- проверка основных гипотез
- разработка начальных моделей.

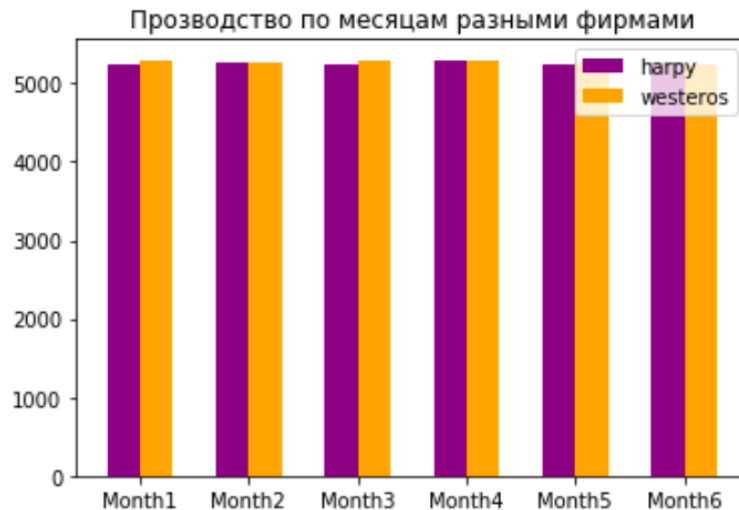
Основные средства разведочного анализа — изучение вероятностных распределений переменных, построение и анализ корреляционных матриц, факторный анализ, дискриминантный анализ, многомерное шкалирование.

Ход работы:

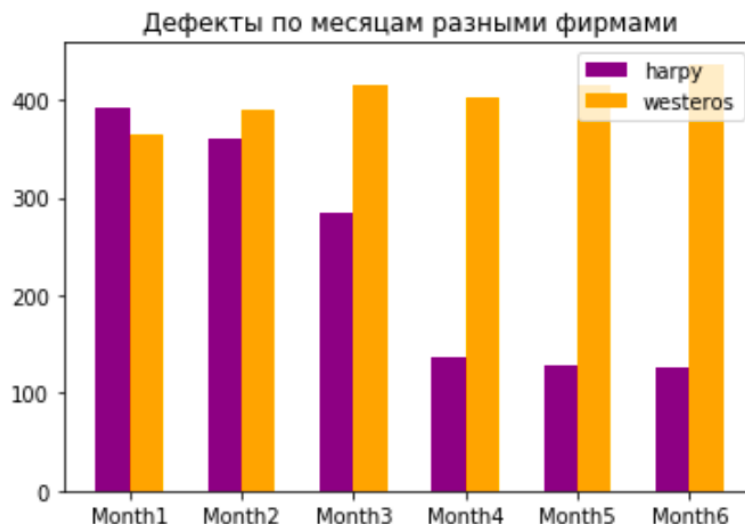
1. Получить данные из CSV таблицы.
2. Провести анализ количества продукции, произведенного компаниями Harry и Westeros каждый месяц.
3. Для большей наглядности изобразим разницу количества дефектов на графике.
4. Рассмотреть плотность распределения усредненных по месяцам дефектов к произведенному товару.
5. Построим графики количества дефектов продукции компании для каждого месяца.

Описание программы.

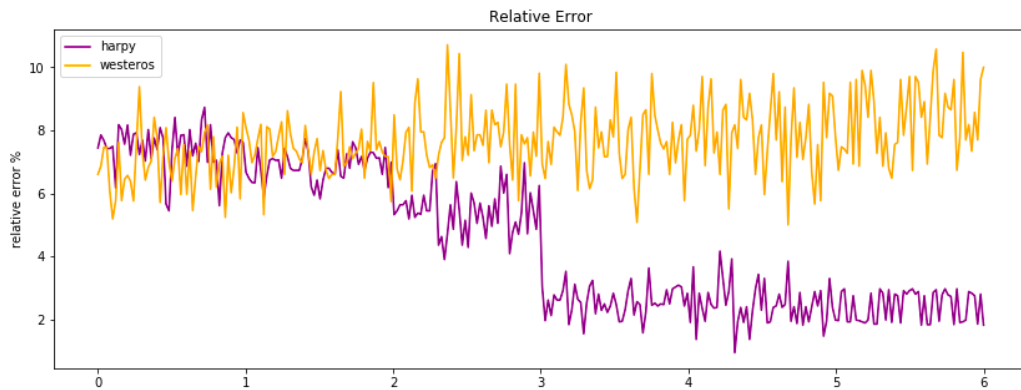
1. Построим диаграмму количества продукции по месяцам. Для этого воспользуемся встроенным инструментом построения bar-гистограмм. Зададим ширину столбцов, цвет. Делаем вывод о том, что компании производят примерно одинаковое количество продукции.



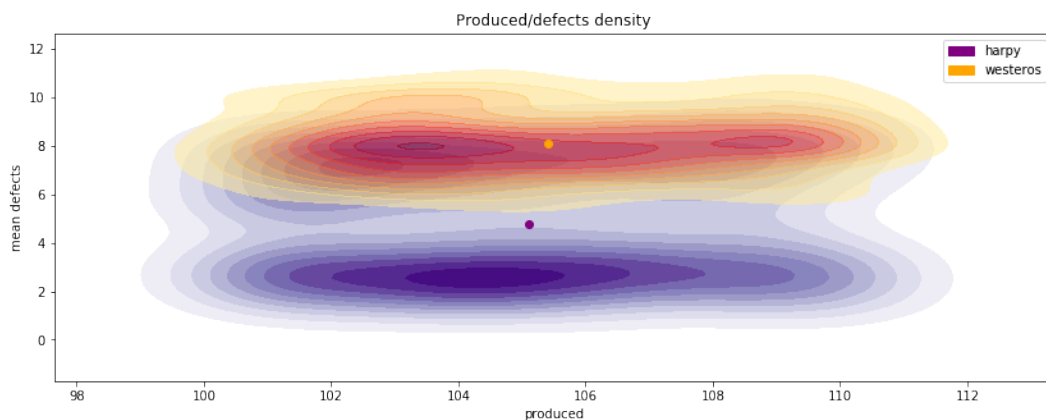
2. Построим диаграмму количества дефектов по месяцам. Заметим сильную разницу в количестве дефектных мечей. Начиная с третьего месяца мечи компании Westeros имеют в разы больше дефектов, чем мечи компании Harpy.



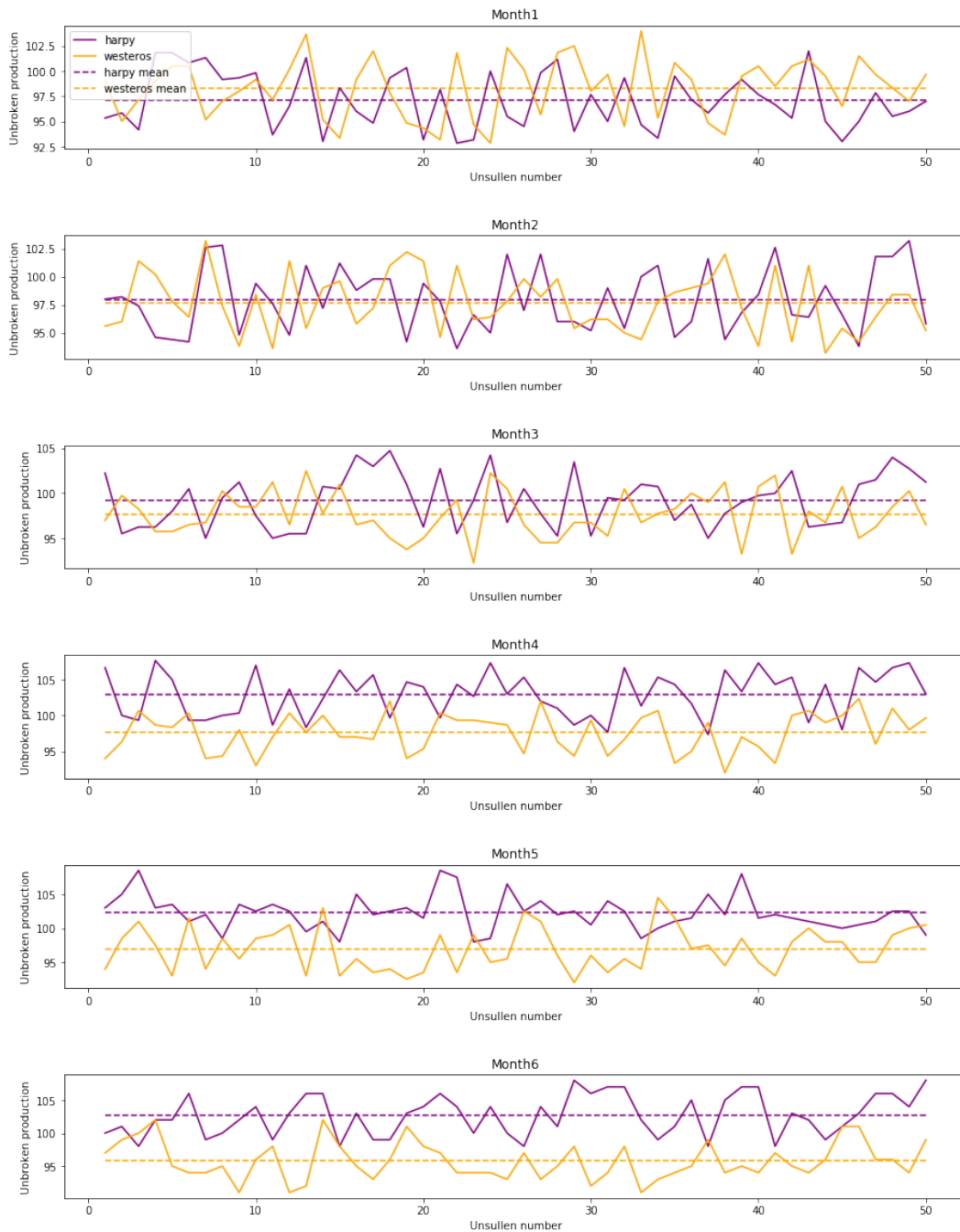
3. Построим график количества дефектов для более наглядной картины. Видим, что с каждым месяцем количество дефектов в мечах компании Harpy уменьшается, причем разница с компанией Westeros значительная.



4. Плотность распределения усредненных по месяцам дефектов к произведенному товару представлена на графике. По этому графику видно что по производству фирмы почти одинаково распределены, а по среднему количеству дефектов Westeros превосходит Harpy. Точки на графике соответствуют среднему значению произведенного товара и средним дефектам за все время. Видно что у Westeros производится в среднем чуть больше, но средние дефекты сильно превышают дефекты Harpy.



5. Построим графики количества дефектов по всем месяцам. Видно, что при примерно одинаковом количестве произведенной продукции среднее количество дефектов компании Harpy ниже, чем компании Westeros. На основании сделанного анализа делаем вывод, что работать стоит с компанией Harpy.



Необходимые компоненты.

Библиотеки:

- **pandas** - библиотека, предназначенная для хранения таблиц, содержит большое количество функций для их обработки.
- **matplotlib.pyplot** – для работы с графиками

Функции:

- **copy** – предоставляет общие операции копирования
- **read_csv(filepath or buffer, sep=',')** – считывает csv файл в dataframe, sep – разделитель
- **groupby()** – группирует элементы таблицы по колонке
- **sum()** – суммирует элементы группы
- **plt.figure(figsize=None).add_subplot(numrows, numcols, fignum)** – **figsize** задаёт размер, **add_subplot** добавляет объект для рисования графика по ук- занным координатам
- **unique()** – возвращает уникальные значения объекта Series
- **fig.plot(x, y, color, label)** – рисует график, где точки имеют координаты (x,y), **label** используется для генерации легенды
- **plt.show()** – показ изображения (графика)
- **plt.subplots(nrows, ncols)** – возвращает объекты Figure и Axes
- **cumsum()** - возвращает совокупную сумму элементов по заданной оси - **where()** - возвращает элементы в зависимости от условия

Программы

- Jupyter Notebook

Вклад участников.

- Липатова Александра - ежемесячное изображение дефектов на графике
- Разумова Вера - составление ReadMe, презентация
- Смирнов Михаил - получение данных из CSV таблицы, анализ количества продукции, произведенного компаниями Harpu и Westeros каждый месяц, изображение количества дефектов на графике