**МИНОБРНАУКИ РОССИИ**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Российский государственный университет туризма и сервиса»**

**ФГБОУ ВО «РГУТИС»**

**Высшая школа сервиса**

наименование структурного подразделения

**Профиль обучения «Информационные сервисы для бизнеса»**

# Отчёт по научно-исследовательской работе

**Исследование и анализ характеристик аккумуляторов разных типов, устанавливаемых на разные автомобили**

Студент

А.Д. Глушенков

(Подпись, дата)

Руководитель НИР

(Подпись, дата)

А.В. Деменев

# Реферат

Отчет 33 с., 26 рис., 4 источн.

АККУМУЛЯТОР, ЕМКОСТЬ, ПУСКОВОЙ ТОК, СОРТИРОВКА, *PYTHON*, БАЗА ДАННЫХ, ПАРАМЕТРЫ, ДИСПЕРСИОННЫЙ АНАЛИЗ.

Объектами исследования являются аккумуляторы колесных машин.

Цель работы – анализ параметров и сравнение аккумуляторов колесных машин при помощи алгоритмов группировки и сортировки данных.

В процессе работы была собрана база данных с параметрами аккумуляторов колесных машин, проведена предобработка данных, а также выполнена сортировка данных при помощи языка программирования *Python*.

В результате исследования выявлена взаимосвязь емкости аккумулятора от массы автомобиля, на который аккумулятор установлен, выявлена взаимосвязь массы аккумулятора от его габаритов, проведен дисперсионный анализ влияния типа аккумуляторов на их характеристики, а также проведен дисперсионный анализ влияния страны-производителя аккумулятора на его емкость.

# Содержание

[Введение 4](#_Toc106304808)

[1 Предобработка данных 5](#_Toc106304809)

[1.1 Принятые допущения 5](#_Toc106304810)

[1.2 Формирование базы данных 5](#_Toc106304811)

[2 Обработка базы данных 7](#_Toc106304812)

[2.1 Группировка по полярности 7](#_Toc106304813)

[2.2 Сортировка по массе автомобиля 8](#_Toc106304814)

[2.3 Сортировка по массе аккумулятора 13](#_Toc106304814)

[3 Дисперсионный анализ 17](#_Toc106304815)

[Заключение 25](#_Toc106304816)

[Список использованных источников 27](#_Toc106304817)

[Приложение 1 28](#_Toc106304818)

[Приложение 2 30](#_Toc106304819)

[Приложение 3 32](#_Toc106304820)

[Приложение 4 33](#_Toc106304821)

# Введение

Автомобильный аккумулятор - тип [электрической аккумуляторной](https://ru.wikipedia.org/wiki/%D0%AD%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%B0%D0%BA%D0%BA%D1%83%D0%BC%D1%83%D0%BB%D1%8F%D1%82%D0%BE%D1%80) [батареи](https://ru.wikipedia.org/wiki/%D0%91%D0%B0%D1%82%D0%B0%D1%80%D0%B5%D1%8F_(%D1%8D%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D1%82%D0%B5%D1%85%D0%BD%D0%B8%D0%BA%D0%B0)), применяемый на автомобильном или мототранспорте. Используется в качестве вспомогательного источника электроэнергии в бортовой сети при неработающем двигателе и для запуска двигателя. Аккумулятор автомобиля содержит химические вещества, которые при взаимодействии производят электрический ток. Два разнородных металла помещаются в кислотную среду, которая называется электролитом. Возникает поток электронов и электроны из одной группы пластин переходят в другую. Обработка, анализ и сортировка данных проведены при помощи языка программирования *Python*, а именно библиотек *Pandas, Matplotlib, Seaborn* и *NumPy*. Данные библиотеки помогают работать с большим объемом информации, сортировать и группировать данные, а также строить графики и тепловые карты.

Цель работы – анализ параметров и сравнение аккумуляторов автомобилей при помощи алгоритмов группировки и сортировки данных.

# Предобработка данных

# Принятые допущения

* + - Для автомобилей в базе данных выбирались аккумуляторы, устанавливаемые при сборке на заводе
    - В качестве массы автомобиля была выбрана снаряженная масса

# Формирование базы данных

База данных содержит информацию об основных параметрах каждого аккумулятора, таких как емкость, Ач; пусковой ток, А; габариты, мм; полярность; напряжение, В; тип, а также применение (на каком автомобиле данный аккумулятор устанавливается), масса автомобиля, кг и страна – производитель данного аккумулятора.

В ходе выполнения работы были собраны необходимые для анализа параметры аккумуляторов [1, 2].

Для работы с базой данных используется программы, код которых был написан на языке программирования *Python* в среде разработки PyCharm с использованием библиотек *NumPy*, *Matplotlib*, *Pandas* и *Seaborn* [3, 4]. Написанные программы позволяют читать таблицы формата .xlsx, визуализировать данные, приведенные в базе данных, сортировать и группировать элементы по тем или иным критериям, создавая при этом новые таблицы формата .xlsx, а также рассчитывать необходимые для дисперсионного анализа параметры.

Полная база данных приведена в Приложении 1.

Часть кода, отвечающая за начало работы с программой, чтение таблицы Excel и выбором дальнейшего действия приведена на рисунке 1.

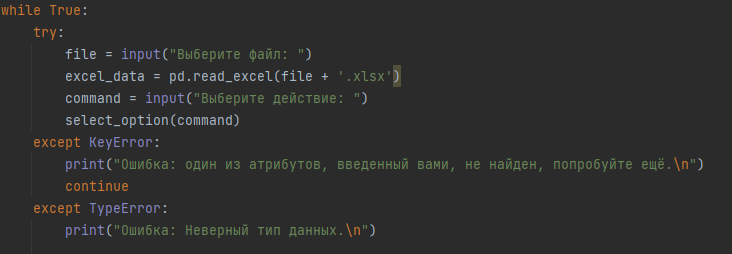


Рисунок 1 – Чтение базы данных и начало работы

# Обработка базы данных

# Группировка по полярности

Выясним в каких автомобилях стоят аккумуляторы с прямой полярностью.

Для этого используем программу, написанную на языке программирования Python. На выходе получим новый Excel - файл с выбранной группой.

Код программы, процесс группировки, а также результат выполнения операций приведены на рисунках 2-4. Полный код программы приведен в Приложении 2.

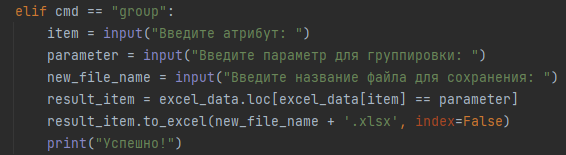


Рисунок 2 – Код программы для группировки базы данных

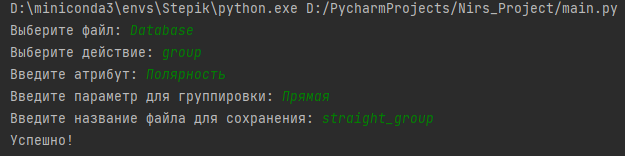


Рисунок 3 – Процесс группировки



Рисунок 4 – Результат группировки аккумуляторов с прямой полярностью

Мы получили новый Excel - файл с результатами (рисунок 4). Из результатов группировки можно сделать вывод, что в нашей базе данных аккумуляторами с прямой полярностью оснащаются автомобили, произведенные в России. Это связано с тем, что прямая полярность была разработана еще для нужд советской автомобильной промышленности. И до сих пор все автомобили, выпускаемые в странах бывшего СССР, комплектуются аккумуляторами с прямой полярностью. Особенность прямой полярности заключается в том, что, если смотреть с лицевой стороны аккумулятора, «плюсовой» вывод установлен слева, а «минусовой» — справа на верхней крышке корпуса АКБ. У аккумуляторных батарей с обратной полярностью, наоборот, «плюс» расположен справа, а «минусовой» вывод – слева.

# Сортировка по массе автомобиля

Проведем сортировку базы данных по массе автомобиля. После чего разобьем ее на несколько групп. Эти действия позволят выявить первичные закономерности, а также подготовят базу данных к дальнейшему дисперсионному анализу.

Для этого также используем программу, написанную на языке программирования Python. На выходе получим новый Excel - файл с отсортированными по массе значениями.

Код программы, процесс сортировки, а также результат выполнения операций приведены на рисунках 5-11.

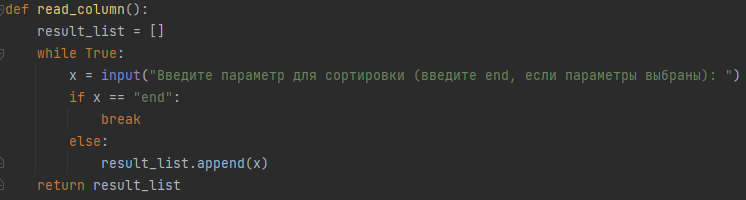


Рисунок 5 – Код программы для сортировки базы данных

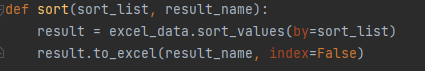


Рисунок 6 – Код программы для сортировки базы данных (продолжение)

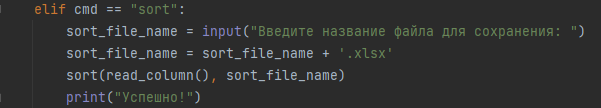


Рисунок 7 – Код программы для сортировки базы данных (продолжение)

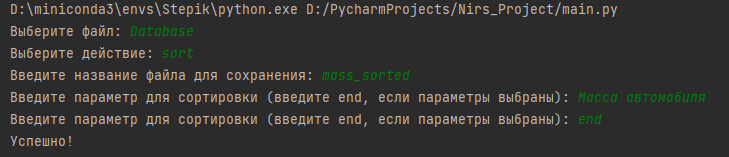


Рисунок 8 – Процесс сортировки

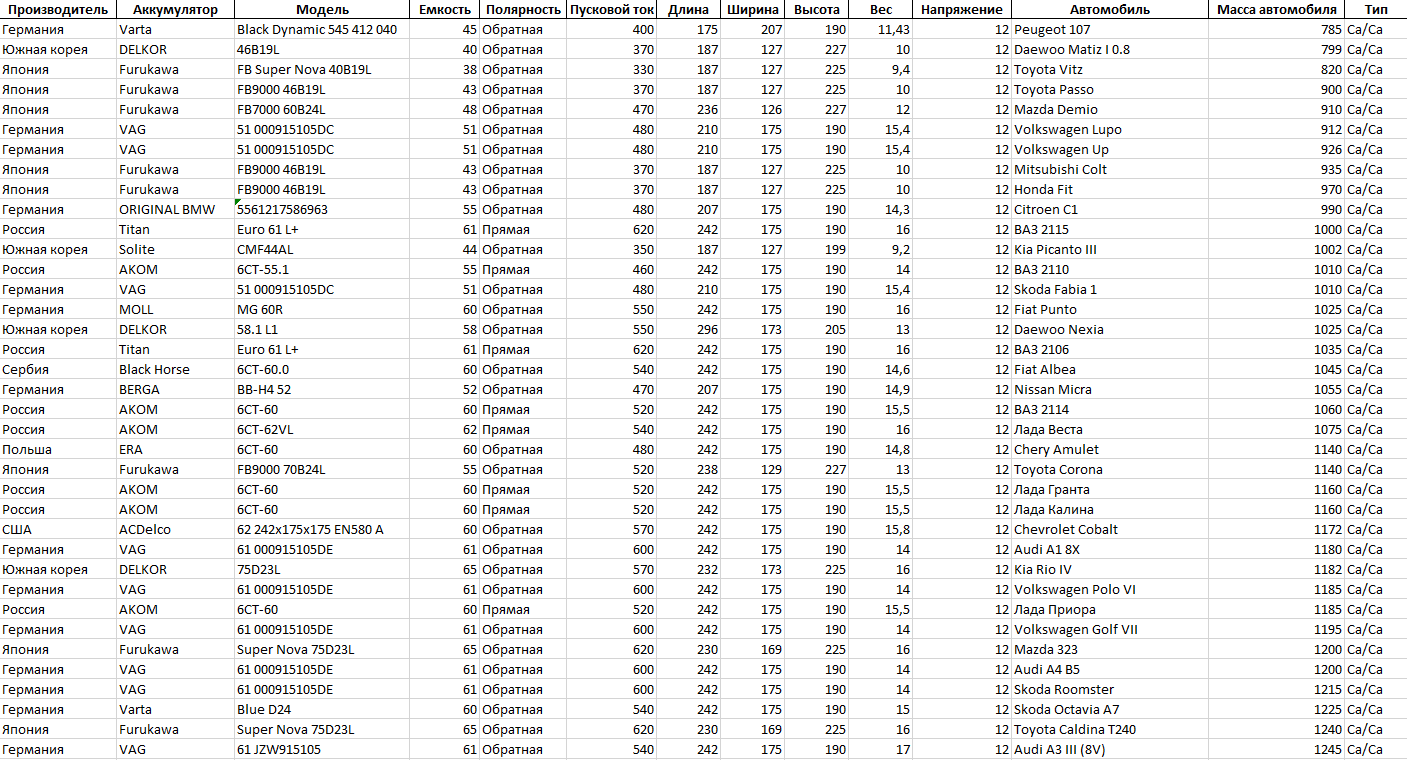


Рисунок 9 – Результат сортировки по массе автомобиля

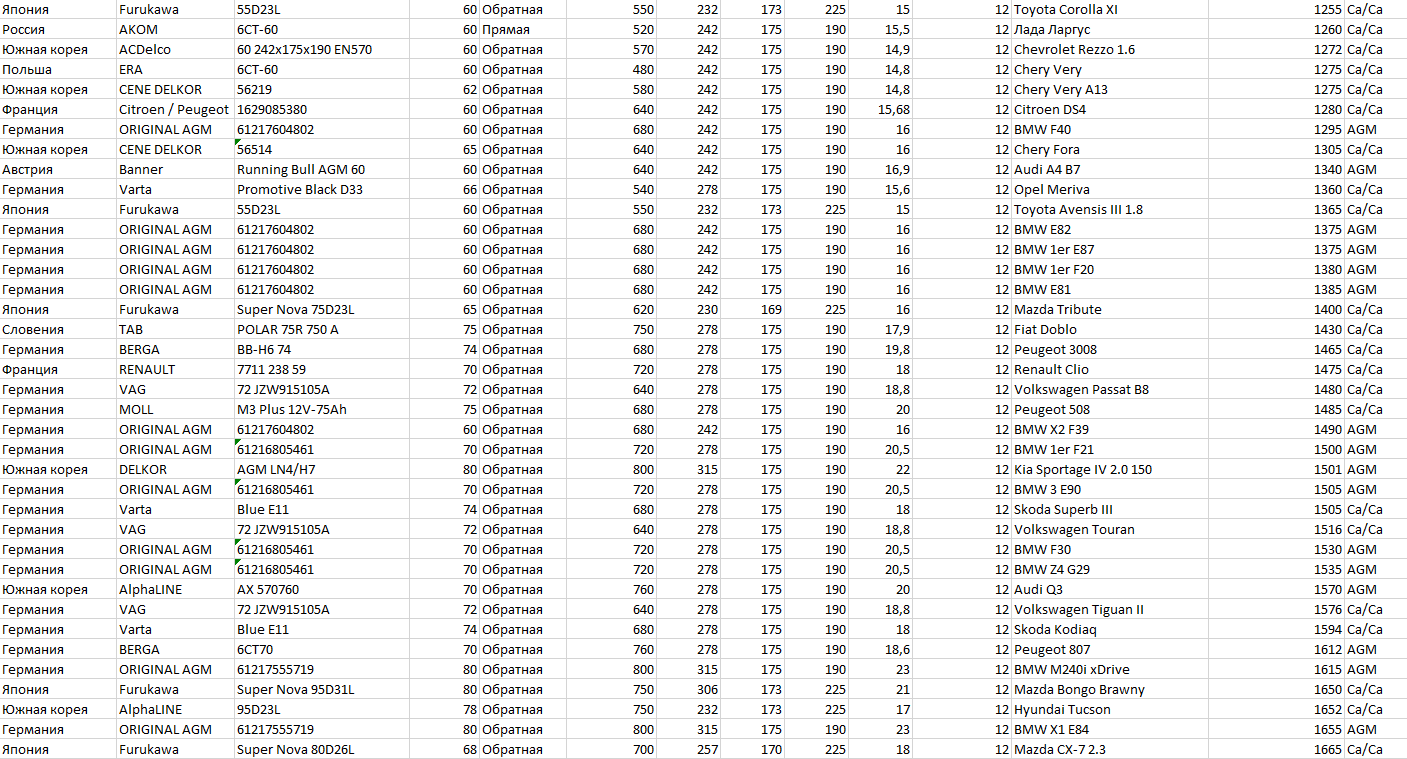


Рисунок 10 – Результат сортировки по массе автомобиля (продолжение)

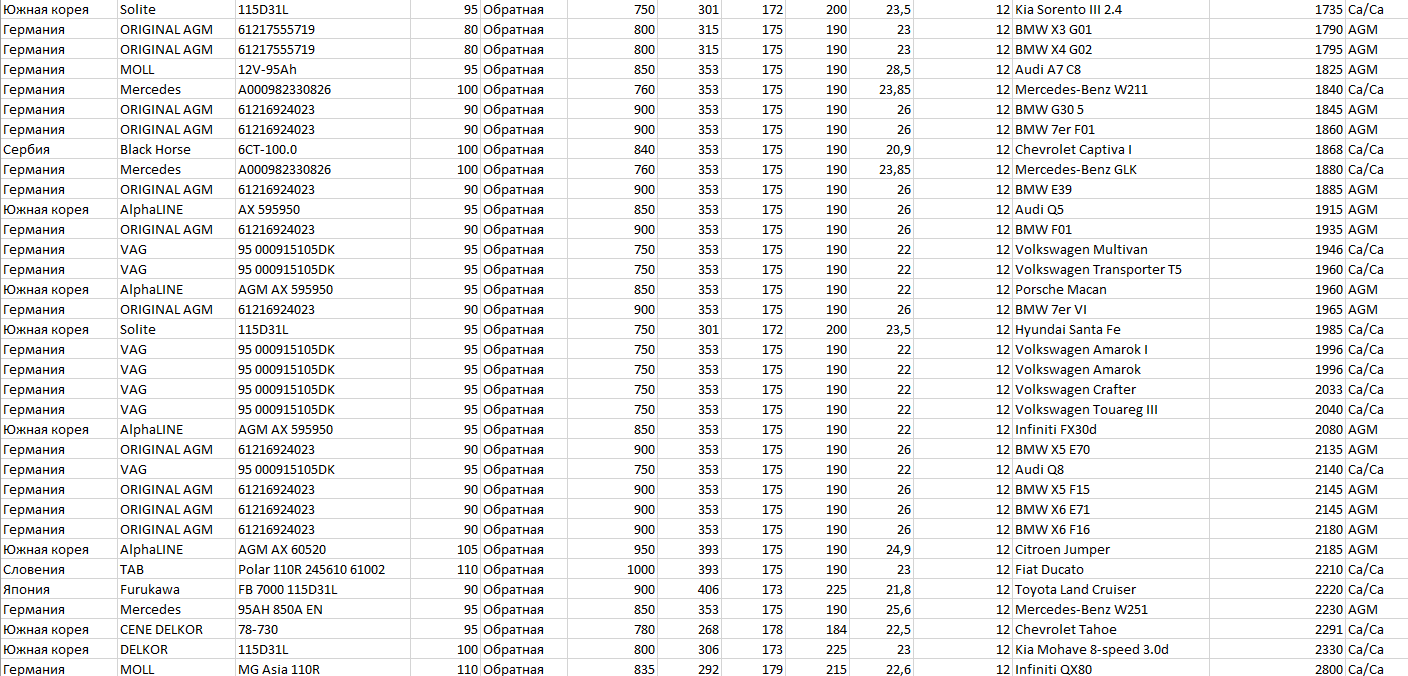


Рисунок 11 – Результат сортировки по массе автомобиля (продолжение)

Мы получили новый Excel - файл с результатами (рисунки 9 - 11). Далее разобьем нашу базу данных на 3 группы по массе автомобилей:

* Группа A – аккумуляторы, устанавливаемые на автомобили с массой до 1200 кг;
* Группа B – аккумуляторы, устанавливаемые на автомобили с массой от 1200 кг до 1600 кг;
* Группа C – аккумуляторы, устанавливаемые на автомобили с массой от 1600 кг.

С помощью программы, написанной на языке программирования Python, построим график средних значений и доверительных интервалов емкостей аккумуляторов для каждой из групп A – C. Полный код программы для построения графика приведен в Приложении 3. График представлен на рисунке 12.

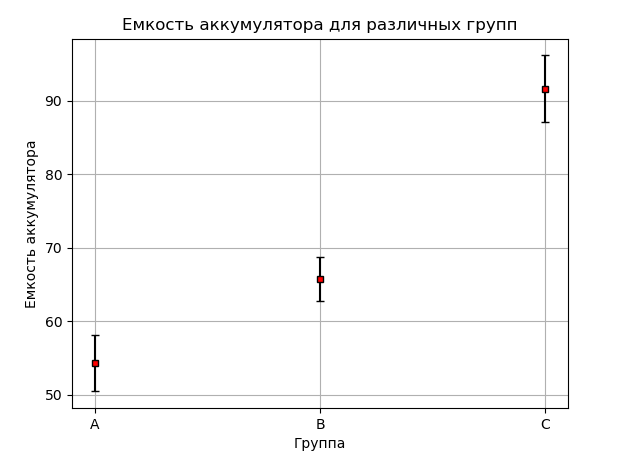


Рисунок 12 – График средних значений и доверительных интервалов емкостей аккумуляторов для каждой из групп A – C

Нам удалось выявить взаимосвязь массы автомобилей с емкостью аккумуляторов, установленных на них. Судя по графику, можно сделать вывод, что с увеличением массы автомобиля емкость установленного на нем аккумулятора также увеличивается. Это можно объяснить тем, что в целом на машины с большей массой устанавливаются более мощные двигатели с большим объемом, которым в свою очередь требуются аккумуляторы с большей мощностью для заведения. Также увеличивается потребление энергии у вспомогательных систем.

# Сортировка по массе аккумулятора

Проведем сортировку базы данных по массе аккумулятора. После чего разобьем ее на несколько групп.

Для этого также используем программу, написанную на языке программирования Python. На выходе получим новый Excel - файл с отсортированными по массе аккумулятора значениями. Также введем такой параметр, как объем аккумулятора. Для его вычисления перемножим габариты аккумулятора.

Код программы сортировки был приведен ранее на рисунках 5-7. Процесс сортировки, а также результат выполнения операций приведены на рисунках 13-16.

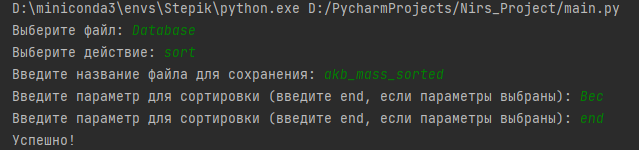


Рисунок 13 – Процесс сортировки

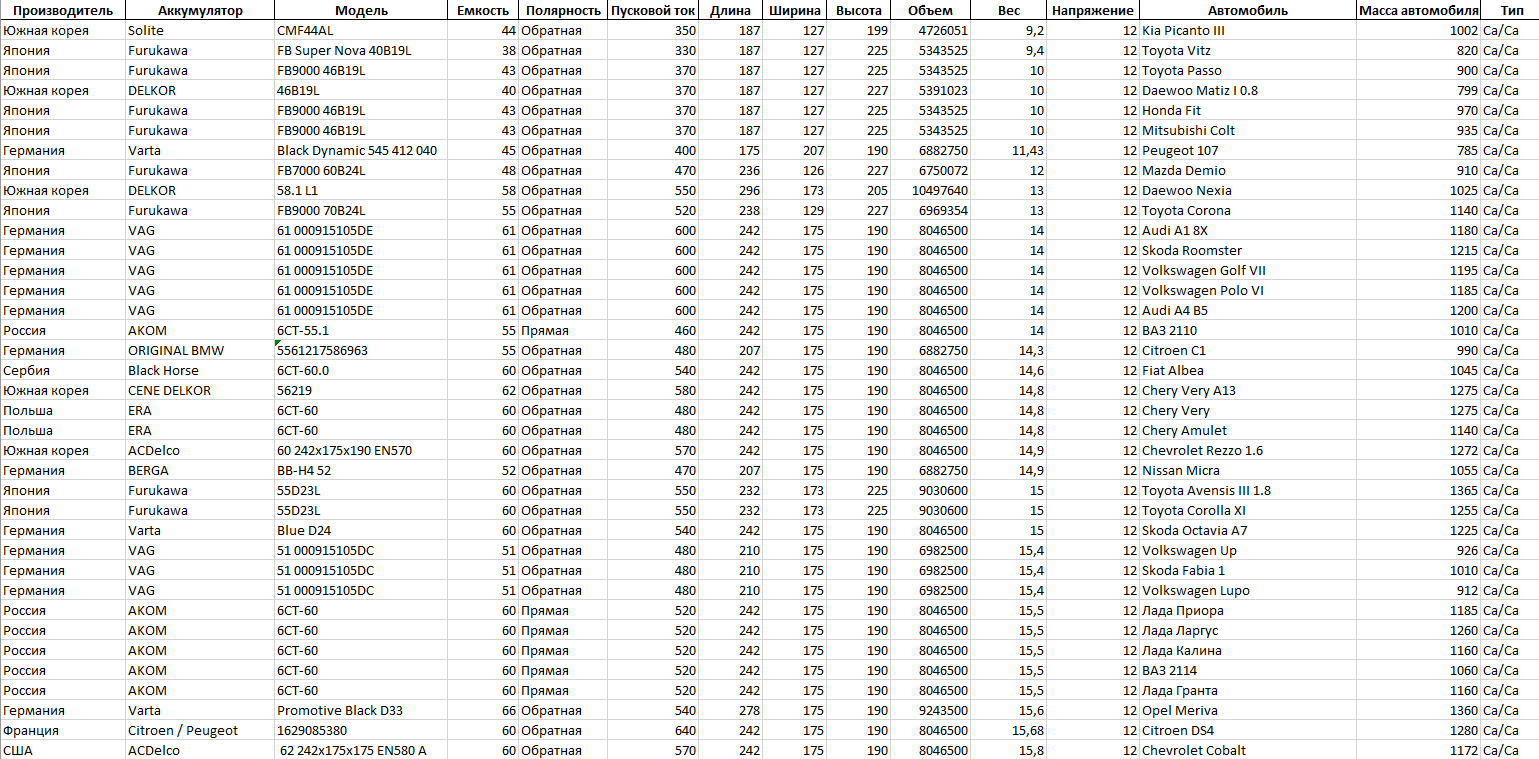


Рисунок 14 – Результат сортировки по массе аккумулятора

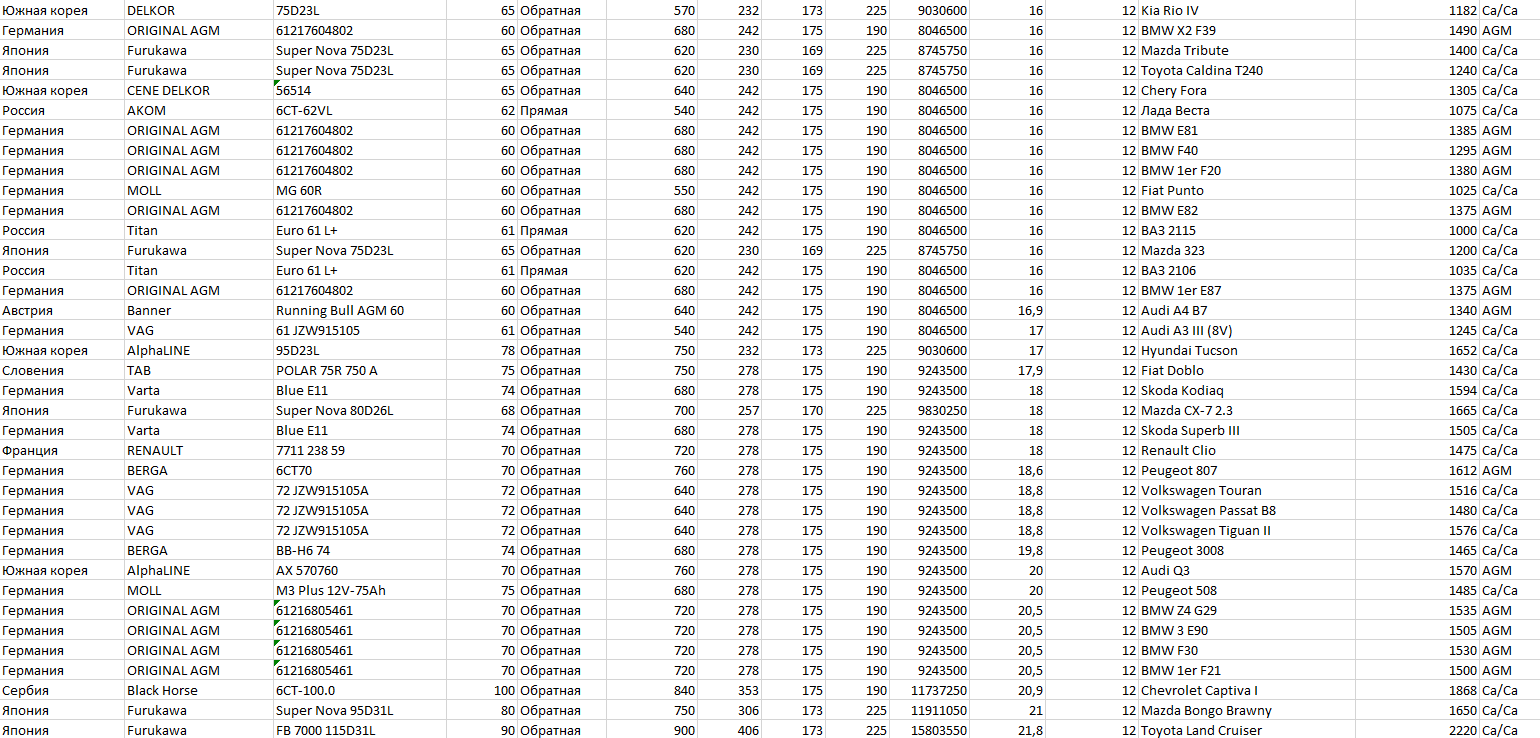


Рисунок 15 – Результат сортировки по массе аккумулятора (продолжение)

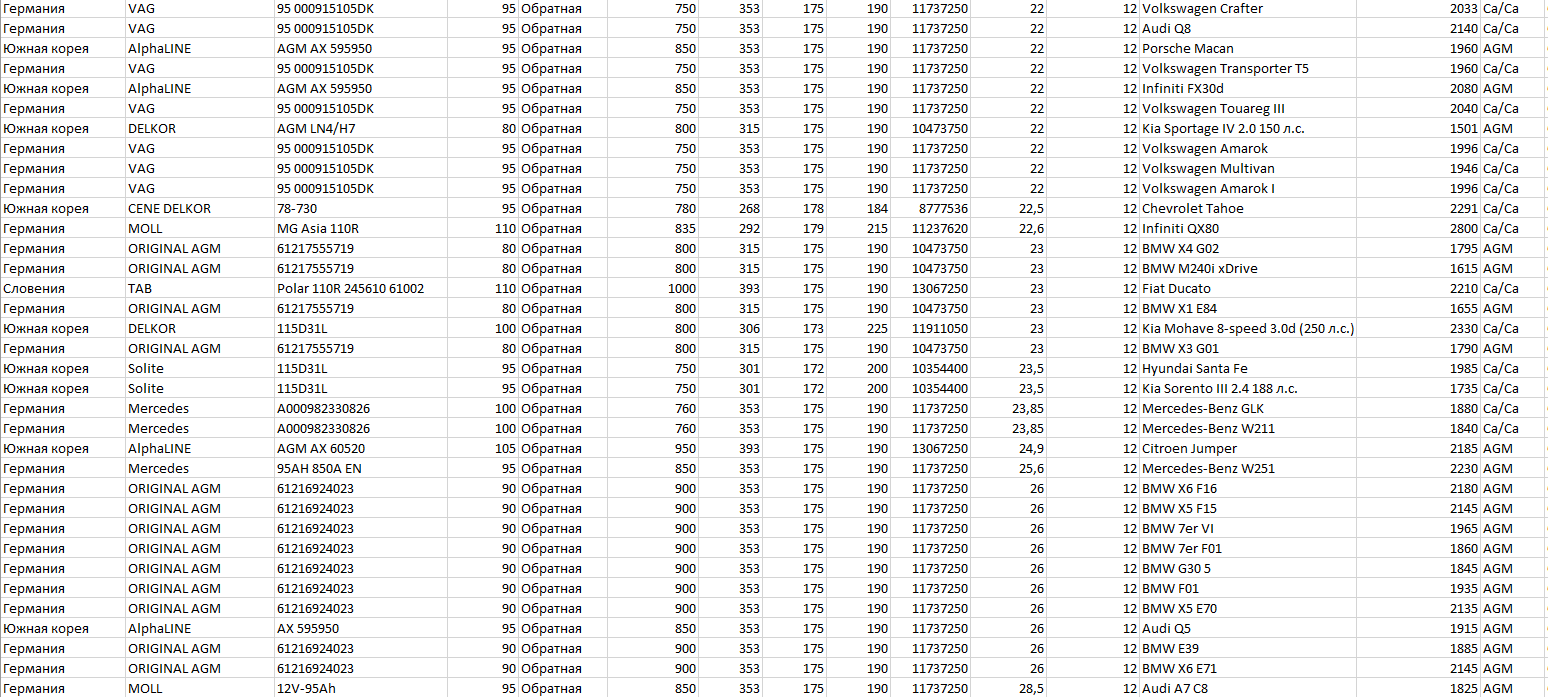


Рисунок 16 – Результат сортировки по массе аккумулятора (продолжение)

Мы получили новый Excel - файл с результатами (рисунки 14 - 16). Далее разобьем нашу базу данных на 3 группы по массе аккумулятора:

* Группа A – аккумуляторы с массой до 16 кг;
* Группа B – аккумуляторы с массой от 16 кг до 22 кг;
* Группа C – аккумуляторы с массой от 22 кг.

Построим график средних значений и доверительных интервалов емкостей аккумуляторов для каждой из групп A – C. График представлен на рисунке 17.

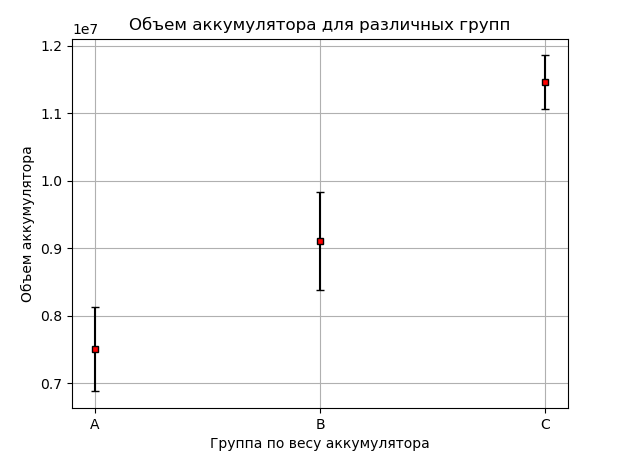


Рисунок 17 – График средних значений и доверительных интервалов объема аккумулятора для каждой из групп A – C

Нам удалось выявить взаимосвязь массы аккумуляторов с их объемом. Судя по графику, можно сделать вывод, что с увеличением габаритов аккумулятора объем аккумулятора также увеличивается.

# Дисперсионный анализ

Выясним влияние некоторых параметров на характеристики аккумуляторов. Для этого подтвердим или опровергнем гипотезу о равенстве средних с помощью дисперсионного анализа, порядок выполнения которого изложен ниже [4].

Выдвигается две гипотезы – первая утверждает равенство средних значений в разных группах; вторая утверждает обратное. В ходе выполнения анализа будет подтверждена одна из них и опровергнута другая.

В начале по формуле (3) находят показатель SST (total sum of squares) – показатель изменчивости исследуемого параметра среди всех групп:

(3)

где – значение параметра конкретного элемента;

– среднее значение параметра среди всех групп.

По формуле (4) находят чисто степеней свободы – число независимых элементов при расчете показателя SST:

(4)

где *N* – число рассматриваемых элементов.

Далее по формуле (5) находят SSW (sum of squares within groups) - показатель изменчивости внутри каждой группы:

(5)

где – среднее значение параметра внутри конкретной группы.

Число степеней свободы для показателя SSW вычисляется по формуле (6):

(6)

где *m* – число групп.

Далее по формуле (7) находят SSB (sum of squares between groups) - показатель межгрупповой изменчивости:

(7)

Число степеней свобод для показателя SSB вычисляется по формуле (8):

(8)

Таким образом, если большая часть изменчивости SST обеспечивается межгрупповой изменчивостью SSB, можно сделать вывод о том, что группы значительно различаются между собой, и наоборот.

Далее рассчитывают основной статистический показатель дисперсионного анализа – F - значение по формуле (9):

(9)

Далее обращаются к статистическому калькулятору, в который вводят полученные значения. Получают значение вероятности большего расхождения между группами. При < 0,05 гипотеза о равенстве средних отклоняется, и наоборот, при > 0,05 гипотеза подтверждается.

Перейдем к реализации данного алгоритма методом программирования на языке *Python.* Рассмотрим влияние типа аккумулятора на пусковой ток. Будем рассматривать автомобили с массой от 1200 кг (группы B и С). Это связано с тем, что в базе данных на автомобили с массой до 1200 кг (группа A) устанавливается только один тип аккумуляторов. Полный код программы, выполняющей дисперсионный анализ, приведен в Приложении 4. Результаты расчета представлены на рисунке 18.

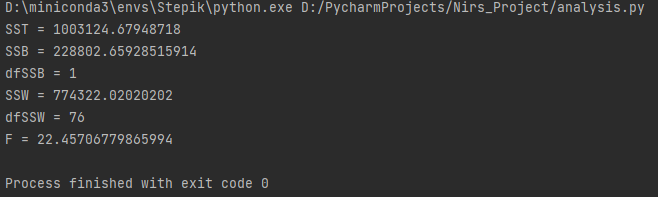


Рисунок 18 – Результаты первого расчета дисперсионного анализа

Из результатов видно, что SSB+SSW=SST, что подтверждает верность расчета.

Обратимся к статистическому калькулятору. Вводя вычисленные значения, получим значение - показателя значимости. Результат представлен на рисунке 19.

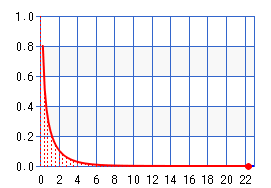


Рисунок 19 – Результат первого расчета калькулятора

Получено значение = 0,00001, что меньше порогового значения 0,05. Согласно критерию, описанному выше, в таком случае гипотеза о равенстве средних опровергается. Это значит, что средние значения пускового тока статистически значимо различаются между группами, то есть пусковой ток зависит от типа аккумулятора. Построим график средних значений и доверительных интервалов. Он изображен на рисунке 20.

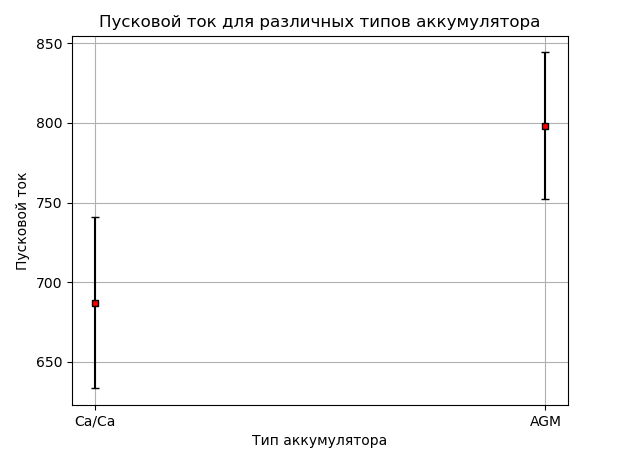


Рисунок 20 – График средних значений и доверительных интервалов пускового тока для разных типов аккумулятора

Удалось выявить статистически значимую взаимосвязь типа аккумулятора с значением пускового тока (F(1,76)=22,45, p<0,05)

Это объясняется тем, что у AGM – аккумуляторов из-за их технологии изготовления внутреннее электрическое сопротивление гораздо меньше, чем у традиционных, и благодаря этому они способны отдавать более высокие токи за короткое время (что особенно важно при холодном пуске двигателя).

Проведем дисперсионный анализ еще раз, для других параметров. Исследуем влияние типа аккумулятора на его емкость. Результаты расчета представлены на рисунке 21.

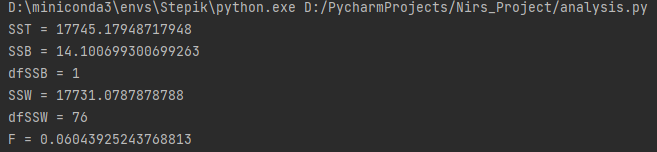


Рисунок 21 – Результаты второго расчета дисперсионного анализа

Аналогично первому расчету по полученным значениям считают -показатель значимости. Результаты представлены на рисунке 22.

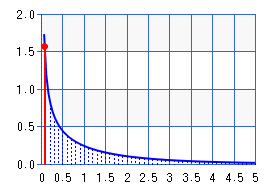


Рисунок 22 – Результат второго расчета калькулятора

Получено значение = 0,807, что выше порогового значения 0,05. Согласно критерию, описанному выше, в таком случае гипотеза о равенстве средних подтверждается. Это значит, что емкость аккумулятора не зависит от его типа. Построим график средних значений и доверительных интервалов. Он изображен на рисунке 23.

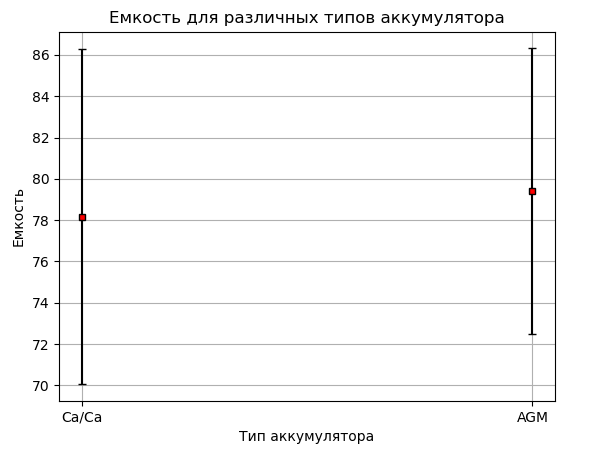


Рисунок 23 – График средних значений и доверительных интервалов пускового тока для разных типов аккумулятора

Не удалось выявить статистически значимую взаимосвязь типа аккумулятора с значением емкости (F(1,76)=0,06, >0,05)

Вновь проведем дисперсионный анализ. Исследуем влияние страны-производителя аккумулятора на его емкость. Ранее, для удобства, мы разбили все страны на европейские и азиатские. Результаты расчета представлены на рисунке 24.

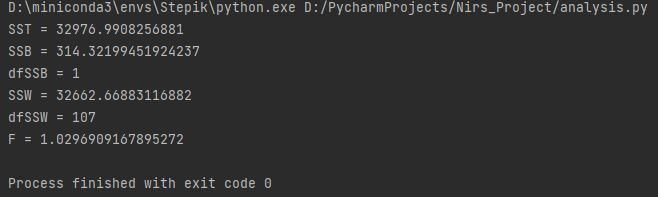


Рисунок 24 – Результаты третьего расчета дисперсионного анализа

Аналогично первому расчету по полученным значениям считают -показатель значимости. Результаты представлены на рисунке 25.

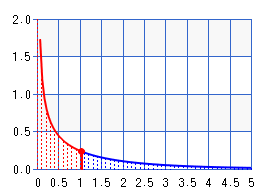


Рисунок 25 – Результат третьего расчета калькулятора

Получено значение = 0,313, что выше порогового значения 0,05. Согласно критерию, описанному выше, в таком случае гипотеза о равенстве средних подтверждается. Это значит, что емкость аккумулятора не зависит от страны-производителя. Построим график средних значений и доверительных интервалов. Он изображен на рисунке 26.

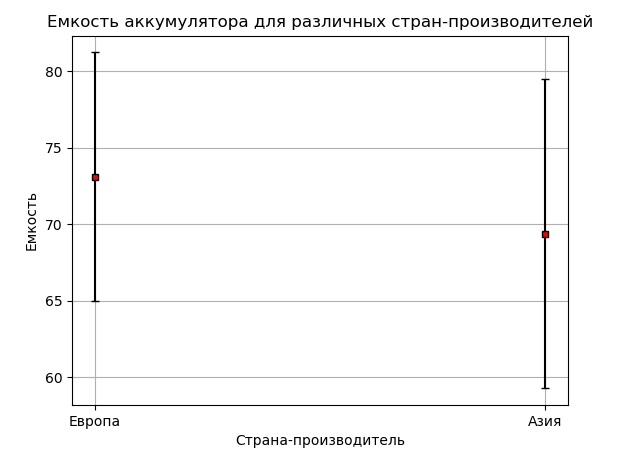


Рисунок 26 - График средних значений и доверительных интервалов емкости для различных стран-производителей

Не удалось выявить статистически значимую взаимосвязь страны-производителя с значением емкости аккумулятора (F(1,107)=1,029, >0,05).

Можно сделать вывод, что как европейские, так и азиатские страны производят одинаковые по емкости аккумуляторы.

# Заключение

В результате выполнения работы был проведен анализ базы данных, содержащей параметры аккумуляторов автомобилей. Были созданы алгоритмы сортировки, группировки, визуализации данных, а также проведения дисперсионного анализа при помощи языка программирования *Python* и его библиотек.

С применением написанного алгоритма группировки были отобраны автомобили с аккумуляторами прямой полярности. С применением написанного алгоритма сортировки были отсортированы аккумуляторы по массе автомобиля, на который они устанавливаются, что позволило в дальнейшем выявить взаимосвязь емкости аккумулятора от массы автомобиля.

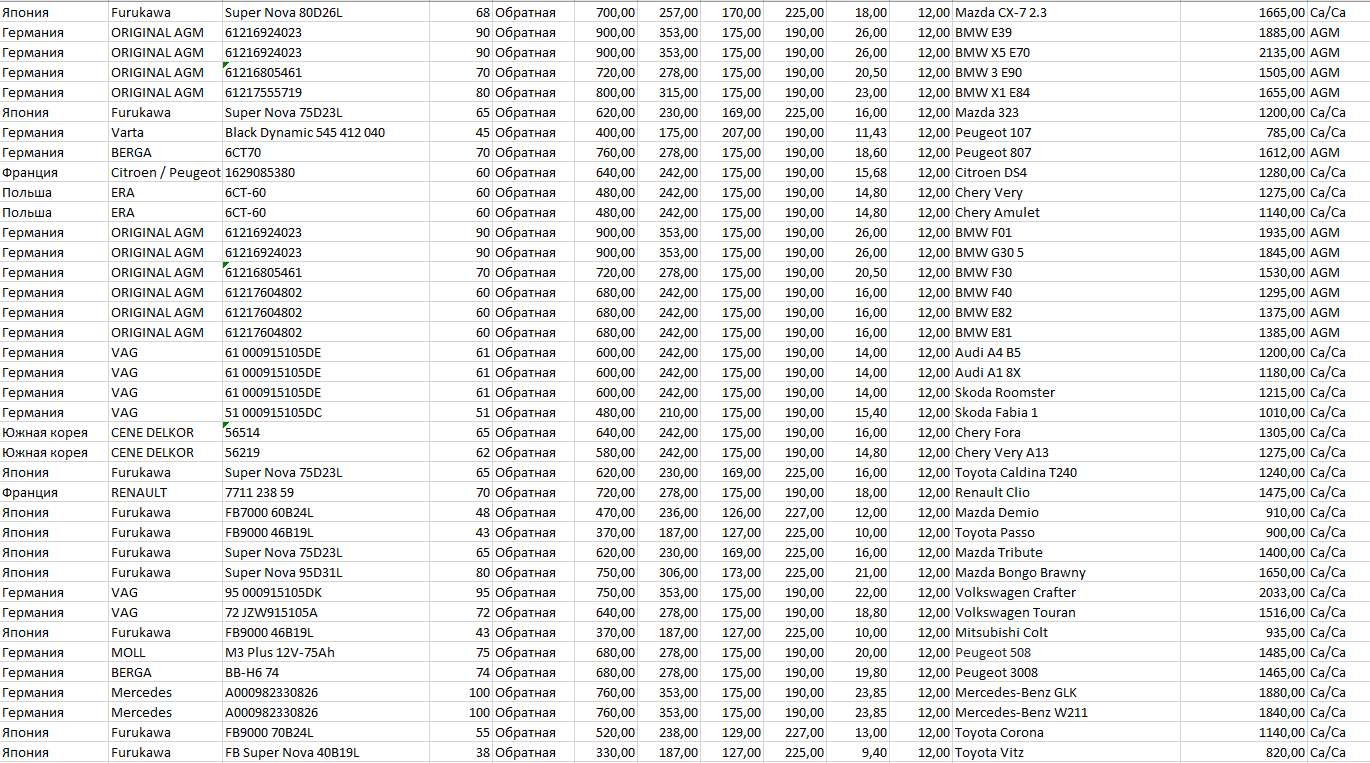
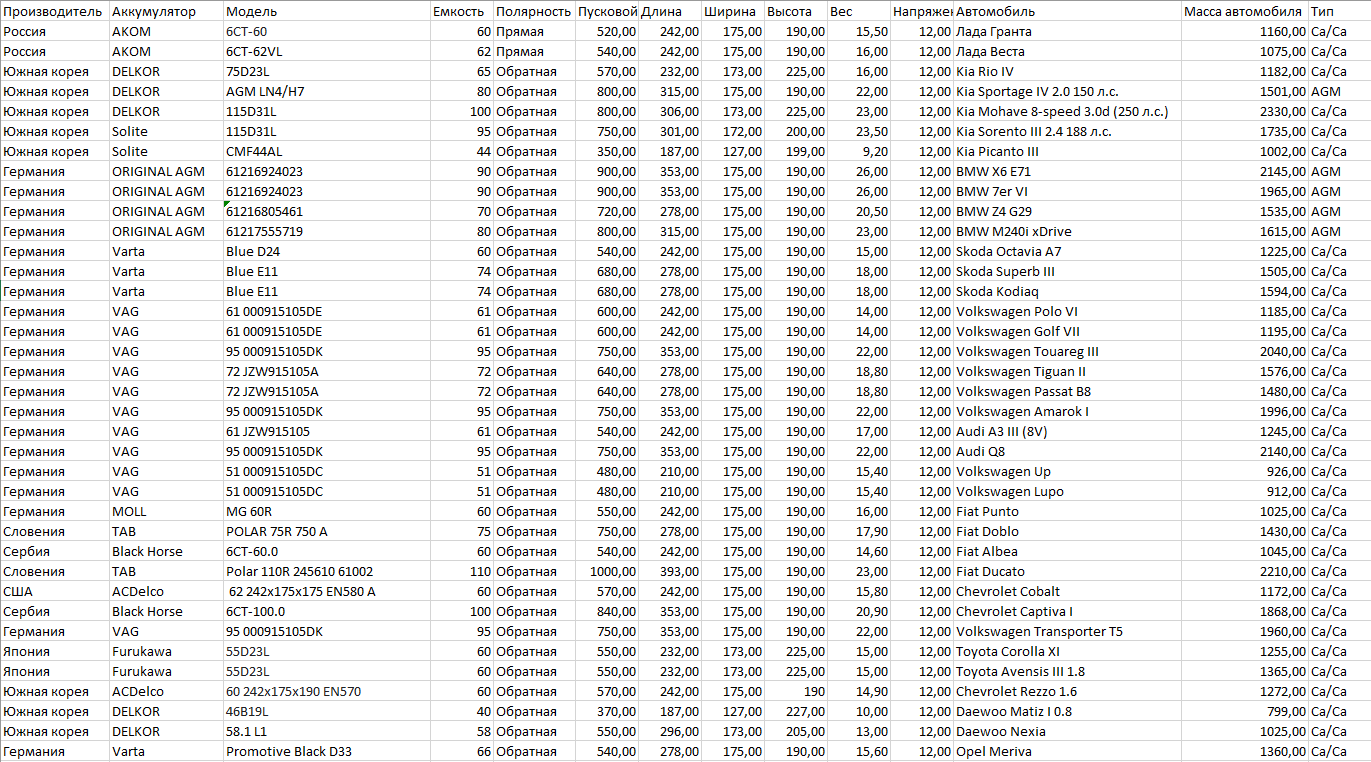
С помощью запрограммированных алгоритмов трижды был проведен дисперсионный анализ – для уточнения величины влияния типа аккумулятора на значение пускового тока, для уточнения величины влияния типа аккумулятора на значение емкости, а также для уточнения величины влияния страны-производителя аккумулятора на значение емкости. В первом случае были получены F-критерий значимости, равный 11,32 и , что позволило опровергнуть гипотезу о равенстве средних и показало, что пусковой ток зависит от типа аккумулятора. Во втором случае были получены F-критерий значимости, равный 0,62 и , что подтвердило гипотезу о равенстве средних и дало вывод об отсутствии значимого влияния типа аккумулятора на значение емкости. Во третьем случае были получены F-критерий значимости, равный 1,03 и , что подтвердило гипотезу о равенстве средних и дало вывод об отсутствии значимого влияния страны-производителя на значение емкости.

По итогу работы были написаны программы, которые можно применять в различных сферах аналитики для практической работы с большими базами данных.

# Список использованных источников

1. Автомобильный аккумулятор // Википедия. [2022]. URL:<https://ru.wikipedia.org/wiki/Автомобильный_аккумулятор>
2. Блог // Магазин аккумуляторов. [2022]. URL: <https://akbmoscow.ru/blog/>
3. Основы Python и анализа данных // Яндекс.Практикум. [2022] URL: <https://practicum.yandex.ru/profile/data-analyst>
4. Основы статистики // Stepik. [2022] URL: <https://stepik.org/course/76/syllabus>

# Приложение 1



# Приложение 2

**import numpy as np**

**import matplotlib.pyplot as plt**

**import pandas as pd**

**import seaborn as sns**

**def read\_column():**

**result\_list = []**

**while True:**

**x = input("Введите параметр для сортировки (введите end, если параметры выбраны): ")**

**if x == "end":**

**break**

**else:**

**result\_list.append(x)**

**return result\_list**

**def heatmap(x, y, z):**

**pivot = excel\_data.pivot\_table(index=[x], columns=[y], values=z, aggfunc=np.average)**

**heatmap = sns.heatmap(pivot)**

**heatmap.set\_title('Correlation Heatmap', fontdict={'fontsize': 18}, pad=12)**

**plt.savefig('heatmap.png', dpi=300, bbox\_inches='tight')**

**def sort(sort\_list, result\_name):**

**result = excel\_data.sort\_values(by=sort\_list)**

**result.to\_excel(result\_name, index=False)**

**def select\_option(cmd):**

**if cmd == "heatmap":**

**first\_value = input("Введите исследуемый элемент: ")**

**second\_value = input("Введите второе значение: ")**

**gradient = input("Градиент распределения: ")**

**heatmap(first\_value, second\_value, gradient)**

**elif cmd == "sort":**

**sort\_file\_name = input("Введите название файла для сохранения: ")**

**sort\_file\_name = sort\_file\_name + '.xlsx'**

**sort(read\_column(), sort\_file\_name)**

**print("Успешно!")**

**elif cmd == "group":**

**item = input("Введите атрибут: ")**

**parameter = input("Введите параметр для группировки: ")**

**group\_file\_name = input("Введите название файла для сохранения: ")**

**result\_item = excel\_data.loc[excel\_data[item] == parameter]**

**result\_item.to\_excel(group\_file\_name + '.xlsx', index=False)**

**print("Успешно!")**

**else:**

**print("Не найдено, попробуйте еще\n")**

**while True:**

**try:**

**file = input("Выберите файл: ")**

**excel\_data = pd.read\_excel(file + '.xlsx')**

**command = input("Выберите действие: ")**

**select\_option(command)**

**except KeyError:**

**print("Ошибка: один из атрибутов, введенный вами, не найден, попробуйте ещё.\n")**

**continue**

**except TypeError:**

**print("Ошибка: Неверный тип данных.\n")**

# Приложение 3

**import pandas as pd**

**import matplotlib.pyplot as plt**

**import statistics**

**data = pd.read\_excel('mass\_sorted.xlsx')**

**grps = pd.unique(data.Группа.values)**

**A = data[data["Группа"] == "A"]["Емкость"]**

**Aavg = sum(A)/len(A)**

**Aerr = statistics.stdev(A)**

**B = data[data["Группа"] == "B"]["Емкость"]**

**Bavg = sum(B)/len(B)**

**Berr = statistics.stdev(B)**

**C = data[data["Группа"] == "C"]["Емкость"]**

**Cavg = sum(C)/len(C)**

**Cerr = statistics.stdev(C)**

**list = ['A', 'B', 'C']**

**listAVG = [Aavg, Bavg, Cavg]**

**listERR = [Aerr/2, Berr/2, Cerr/2]**

**print(list, listAVG)**

**plt.errorbar(x=list, y=listAVG, yerr=listERR, color="black", capsize=3, marker="s", markersize=5, mfc="red", mec="black", fmt='o')**

**plt.title('Емкость аккумулятора для различных групп')**

**plt.grid()**

**plt.xlabel('Группа')**

**plt.ylabel('Емкость аккумулятора')**

**plt.show()**

# Приложение 4

**import pandas as pd**

**excel\_data\_df = pd.read\_excel('disp.xlsx')**

**indep = excel\_data\_df["Тип"]**

**dep = excel\_data\_df["Емкость"]**

**N = 0**

**group\_count = {}**

**group\_result = {}**

**res = 0**

**counter = 0**

**for item in dep:**

**counter = counter + 1**

**res = res + item**

**X = res / counter**

**SST = 0**

**SSW = 0**

**SSB = 0**

**for item in dep:**

**prom = (item - X) \*\* 2**

**SST = SST + prom**

**for item in indep:**

**if item not in group\_count:**

**group\_count[item] = 1**

**else:**

**group\_count[item] = group\_count[item] + 1**

**for item in group\_count:**

**N = N + group\_count[item]**

**result\_item = excel\_data\_df.loc[excel\_data\_df["Тип"] == item]**

**group\_result[item] = sum(result\_item["Емкость"]) / group\_count[item]**

**for item in group\_result:**

**result\_item = excel\_data\_df.loc[excel\_data\_df["Тип"] == item]**

**for i in result\_item["Емкость"]:**

**SSW = SSW + ((i - group\_result[item]) \*\* 2)**

**dFSSW = N - len(group\_result)**

**for item in group\_result:**

**SSB = SSB + group\_count[item] \* ((group\_result[item] - X) \*\* 2)**

**dFSSB = len(group\_result) - 1**

**F = (SSB / dFSSB) / (SSW / dFSSW)**

**print('SST =',SST)**

**print('SSB =',SSB)**

**print('dfSSB =', dFSSB)**

**print('SSW =',SSW)**

**print('dfSSW =', dFSSW)**

**print('F =',F)**