



## Направление «Аналитик данных» (Data Analyst), МТС

### Аннотация

Добро пожаловать на платформу карьерного тест-драйва **Shift + Enter**! Здесь ты можешь решить задания от МТС, компании, где работа построена на сплоченных продуктовых командах, использующих agile<sup>1</sup> методологию. Департамент Big Data МТС – место, где телеком-данные превращаются в востребованные IT-продукты.

Не уппусти шанс попробовать себя в роли начинающего аналитика данных в Big Data МТС, это позволит тебе:

- Погрузиться в работу над реальными проектами департамента в рамках твоего стажерского спринта.
- Улучшить знания SQL и основных библиотек Python для работы с данными.

### Развиваемые компетенции

По результатам выполнения заданий ты сможешь развить такие навыки, как:

- Подготовка и визуализация данных.
- Анализ больших данных, хранящихся в разных таблицах.

А также научишься искать и структурировать информацию из открытых источников, прокачаешь логику и анализ, сможешь проявить аккуратность и внимательность к деталям.

### Описание подзадач

Команда Big Data МТС создала несколько десятков аналитических сервисов. Самые успешные из них уже стали элементами экосистемы МТС. Настало твое время – ты готов виртуально присоединиться к работе департамента и помочь с решением задач, связанных с анализом и визуализацией данных?

### Рекомендуемый тайминг

Выполнение всего блока заданий займет у тебя не более 80–100 минут.

1. 20–30 минут на первое задание.
2. 30 минут на второе задание.
3. 30–40 минут на третье задание.

### Информация о загрузке решения

Направление «Аналитик данных» содержит несколько подзадач. Можно загрузить файл, содержащий решение части заданий, но по возможности постарайся сделать их все.

Желаем удачи!

---

<sup>1</sup> Agile – это группа методик для гибкого управления проектами в команде разработки. Рабочий процесс при таком подходе разбивается на небольшие временные промежутки, их еще называют спринтами или итерациями. Во время каждого спринта команда разработки создает часть продукта, которую можно протестировать и оценить.



## Задание 1. Составь SQL-запрос, помоги коллегам

Антон<sup>2</sup>, руководитель группы рекламной аналитики центра MTS Big Data попросил тебя подключиться к первой задаче и помочь команде сервиса «[МТС Маркетолог](#)».

Привет!

Уверен, что ты в курсе новостей об умном сервисе «МТС Маркетолог», который позволяет нашим клиентам самостоятельно продвигать свои товары и услуги посредством таргетинга и ресурсов Big Data MTC с машинным обучением.

Эту платформа была создана для самозанятых, фрилансеров, а также представителей малого и среднего бизнеса. Например, ее планирует использовать индивидуальный предприниматель, только что открывший кофейню по адресу г. Тверь, ул. Трёхсвятская, д.3, который хочет привлечь как можно больше новых посетителей из числа абонентов МТС.

Твоя задача – использовать данные, которые у нас имеются, чтобы:

- Описать алгоритм, который позволит ему собрать сегмент абонентов.
- Предложить наиболее подходящий способ привлечения новых клиентов.

Пока что ты работаешь с фрагментами (сэмплами) данных. Но я рекомендую описать выгрузку сегмента в виде SQL-запроса, так как это позволит тебе применить уже готовый запрос позднее, когда я пришлю тебе весь массив данных.

Hints. Описание таблиц с данными ты найдешь во вложении, а ссылку для скачивания данных – в разделе «Полезные материалы».

Спасибо!

### Полезные материалы

- [Сэмплы данных](#) в формате .csv (rar-архив).
- [Руководство](#) по анализу данных с SQL.

### Формат конечного результата

Файл в формате .docx, содержащий описание алгоритма, предложения по привлечению и SQL-запрос для выгрузки сегмента абонентов.

### Форма загрузки результата

Пожалуйста, загрузи свой вариант ответа в формате zip-архива, используя инструмент «Загрузить решение». Необходимо сформировать единый zip-архив, содержащий решение одного или всех заданий по выбранной специальности.

### Пример решения

У тебя будет возможность ознакомиться с примером решения задания от эксперта после отправки собственной версии.

---

<sup>2</sup> Все имена и названия вымышленные, любые совпадения случайны. Данные заданий могут быть изменены в целях конфиденциальности.

## Вложение 1 – Информация о таблицах, содержащих данные

- **cnum\_d:** таблица со звонками всех абонентов МТС по всей России cnum\_d с полями: msisdn (номер абонента), ts (дата и время звонка), called\_party\_number (номер, с которого или на который звонил абонент, в случае sms в этом поле может быть название компании-отправителя или короткий номер), call\_type (тип вызова sms – 2 или звонок – 1), direction (направление вызова – входящий (I) или исходящий (O))

cnum\_d

msisdn	ts	called_party_number	call_type	direction
...	...	...	...	...

- **host\_d:** таблица с данными посещения хостов всех абонентов МТС по всей России host\_id с полями: msisdn (номер абонента), ts (дата и время посещения), host (URL сайта типа yandex.ru)

host\_d

msisdn	ts	host
...	...	...

- **geo\_d:** таблица с данными, обновляемыми раз в 10 минут о местонахождении абонента geo\_d с полями: msisdn (номер абонента), ts (дата и время определения местонахождения), geo\_area\_key (идентификатор зоны местонахождения абонента)

geo\_d

msisdn	ts	geo_area_key
...	...	...

- **twogis\_phones:** справочник 2GIS, в котором есть рубрика, к которой относится номер телефона организаций и название самой организации, т.е. таблица twogis с полями: rubric, phone\_num, organization\_name

twogis\_phones

rubric	phone_num	organization_name
...	...	...

- **twogis\_hosts:** справочник 2GIS, в котором есть рубрика, к которой относится сайт организации и название самой организации, т.е. таблица twogis с полями: rubric, host\_name, organization\_name

twogis\_hosts

rubric	host_name	organization_name
...	...	...

- **cities\_regions**: справочник с идентификаторами зон, которые соответствуют региону и городу, району города в которых они находятся, т.е. таблица cities\_regions с полями: geo\_area\_key, city\_name, region, city\_area

**cities\_regions**

geo_area_key	city_name	region	city_area
...	...	...	...

- набор оборудования для рассылки sms сообщений любому абоненту МТС с любым текстом, а также call-центр для обзвона абонентов.



## Задание 2. Построй диаграммы и графики для визуализации зависимостей

Пришло время поработать над вторым проектом. Открыв почту, ты внимательно изучаешь детали задания, которые прислал Антон.

Привет,

Благодарю за помощь с сегментацией абонентов, отличная работа!

У нас есть файл с данными о покупках, совершенных в трех ресторанах быстрого питания в течение февраля 2023 года. Тебе необходимо построить несколько визуализаций, которые помогли бы ответить на следующие вопросы:

1. Какова динамика трат посетителей в ресторанах? Есть ли дни наибольшего спроса? Можно ли проследить тренд? Какой ресторан находится на втором месте по популярности?
2. Каково распределение покупателей по возрасту в каждом из ресторанов?
3. Как можно визуализировать одновременно 3 параметра: сумму покупок, средний чек и количество покупателей?

Hints. При работе над этим заданием, можешь использовать любой доступный инструмент анализа, например Excel, R или Python (лично я бы решал задачу, используя Excel).

Жду твое решение, спасибо!

### Полезные материалы

[Файл с исходными данными](#) о покупках, совершенных в трех ресторанах быстрого питания.

### Формат конечного результата

Файл с нужными диаграммами / матрицами и ответами на вопросы.

### Форма загрузки результата

Пожалуйста, загрузи свой вариант ответа в формате zip-архива, используя инструмент «Загрузить решение». Необходимо сформировать единый zip-архив, содержащий решение одного или всех заданий по выбранной специальности.

### Пример решения

У тебя будет возможность ознакомиться с примером решения задания от эксперта после отправки собственной версии.



## Задание 3. Составь SQL-запрос, помоги найти инсайты в данных

Ты отлично справился с первыми двумя заданиями и готов приступить к финальному. Чуть позже ты увидел сообщение от Антона с вводными по этому проекту.

Привет,

Твой заключительный проект, как и первый, вновь связан с анализом данных и написанием SQL-запроса для выбора нужной информации.

У нас есть несколько датасетов, включая таблицу с историей владения мобильными телефонами, содержащую информацию о владельцах (сотрудниках одной компании), вне зависимости от того, у какого оператора связи они находились или находятся сейчас. Во второй таблице есть данные о времени пребывания этих людей с определенными операторами связи.

Используй имеющиеся данные, чтобы установить максимальное время владения телефоном для выбранного оператора связи. Эта информация позволит нам сделать выводы и сформулировать начальные гипотезы о том, существует ли какая-то связь между марками телефонов и длительностью пребывания их владельцев у определенных операторов сотовой связи.

Пока что ты снова работаешь с фрагментами данных. Но если ты напишешь SQL-запрос сейчас, то это позволит нам использовать его позднее, когда мы выгрузим весь массив данных.

Hints. Описание таблиц с данными ты найдешь во вложении, а ссылку для скачивания данных – в разделе «Полезные материалы».

Пожалуйста, пришли свое решение с пояснениями до конца рабочего дня. Большое спасибо!

### Полезные материалы

[Сэмплы данных](#) в формате .csv (zip-архив).

### Формат конечного результата

Файл в формате .docx, содержащий SQL-запрос и пояснения к нему.

### Форма загрузки результата

Пожалуйста, загрузи свой вариант ответа в формате zip-архива, используя инструмент «Загрузить решение». Необходимо сформировать единый zip-архив, содержащий решение одного или всех заданий по выбранной специальности.

### Пример решения

У тебя будет возможность ознакомиться с примером решения задания от эксперта после отправки собственной версии.



## Вложение 2 – Информация о таблицах, содержащих данные

- **devices:** таблица с данными о девайсах с полями:

devices.csv	empl_id	date_from	date_to	device
	ID сотрудника	дата начала пользования устройством	дата конца пользования устройством	марка устройства

- **emp\_name:** таблица с данными о сотрудниках с полями:

emp_name.csv	empl_id	first_name
	ID сотрудника	имя сотрудника

- **operators:** таблица с данными об операторе с полями:

operators.csv	empl_id	date_from	date_to	op_name
	ID сотрудника	дата подключения к оператору	дата отключения от оператора	название оператора