1. **Паспорт Образовательной программы**

**«Аналитик данных: основы профессии»**

|  |  |
| --- | --- |
| **Версия программы** | **1** |
| **Дата Версии** | 09.10.2020 |

1. **Сведения о Провайдере**

|  |  |  |
| --- | --- | --- |
| 1.1 | Провайдер | Общество с ограниченной ответственностью «Центр онлайн-обучения Нетология-групп» |
| 1.2 | Логотип образовательной организации | <https://drive.google.com/file/d/1tbT_26-AiHUALMw1qe74H6MJySCBAyBJ/view?usp=sharing> |
| 1.3 | Провайдер ИНН | 7715964180 |
| 1.4 | Ответственный за программу ФИО | Герасимова Елена |
| 1.5 | Ответственный должность | Руководитель БЮ Аналитика и Data Science |
| 1.6 | Ответственный Телефон | +79067563636 |
| 1.7 | Ответственный Е-mail | gerasimova@netology.ru |

1. **Основные Данные**

|  |  |  |
| --- | --- | --- |
| **№** | **Название** | **Описание** |
| 2.1 | Название программы | Аналитик данных: основы профессии |
| 2.2 | Ссылка на страницу программы | https://netology.ru/programs/da-basic#/ |
| 2.3 | Формат обучения | Онлайн |
|  | Подтверждение от ОО наличия возможности реализации образовательной программы с применением электронного обучения и (или) дистанционных образовательных технологий с возможностью передачи данных в форме элементов цифрового следа | да, подтверждаем |
| 2.4 | Уровень сложности | Начальный |
| 2.5 | Количество академических часов | 72 |
|  | Практикоориентированный характер образовательной программы: не менее 50 % трудоёмкости учебной деятельности отведено практическим занятиям и (или) выполнению практических заданий в режиме самостоятельной работы (кол-во академических часов) | 34 часа - теоретические занятия  38 часов - практические занятия |
| 2.6 | Стоимость обучения одного обучающегося по образовательной программе, а также предоставление ссылок на 3 (три) аналогичные образовательные программы иных организаций, осуществляющих обучение, для оценки объективности стоимости или обоснование уникальности представленной образовательной программы в случае отсутствия аналогичных образовательных программ на рынке образовательных услуг | 30 000  Data Analysis with Python and SQL  <https://www.udacity.com/course/data-analyst-nanodegree--nd002>  28 300, 19 часов  Advanced Data Analysis  <https://www.udacity.com/course/data-analyst-nanodegree--nd002>  40 000, 19 часов  Аналитики Big Data  <https://praktikum.yandex.ru/data-analyst?utm_source=google&utm_medium=cpc&utm_campaign=Google_Search_Data-analyst_Obuchenie-2&utm_content=adapt&utm_term=%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D1%82%D0%B8%D0%BA%20%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85%20%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5&gclid=CjwKCAjw5Kv7BRBSEiwAXGDElXbcQFXHc4hmgyVMdNE1B6892ma12hJIa1d8RMfACcgJsYVHULAsLhoCMZoQAvD_BwE>  65 000, |
| 2.7 | Минимальное количество человек на курсе | 10 |
| 2.8 | Максимальное количество человек на курсе | 10000 |
| 2.9 | Данные о количестве слушателей, ранее успешно прошедших обучение по образовательной программе | 2000 |
| 2.10 | Формы аттестации | зачет |
|  | Указание на область реализации компетенций цифровой экономики, к которой в большей степени относится образовательная программа, в соответствии с Перечнем областей | Большие данные |

1. **Аннотация программы**

Программа составлена в соответствии с Перечнем ключевых компетенций цифровой экономики утвержденным П[риказом Минэкономразвития России от 24.01.2020 № 41](https://rulaws.ru/acts/Prikaz-Minekonomrazvitiya-Rossii-ot-24.01.2020-N-41/) и направлена на освоение навыков писать простые SQL запросы, чтобы получать данные из базы данных; писать джоины, чтобы создавать новые таблицы; группировать, фильтровать данные; импортировать и экспортировать данные; использовать аналитические функции SQL, чтобы изучать характеристики данных. В результате обучения по программе слушатель получит представление о профессии& В планируемых результатах новых компетенций слушатели смогут: сформулировать задачи, подбирать алгоритмы и метрики под них, собирать данные из баз данных с помощью SQL, визуализировать данные, с помощью python-библиотек находить закономерности в данных, очищать и предобрабатывать данные, выбирать и создавать фичи для построения моделей, строить модели машинного обучения с помощью библиотек sklearn, оценивать качество этих моделей, интерпретировать результаты и составлять отчеты об исследовании. Для закрепления материала предусмотрены практические задания. Успешным прохождением обучения считается выполнение итоговых заданий по каждому модулю.

Общество с ограниченной ответственностью «Центр онлайн-обучения Нетология-групп»

Дополнительная профессиональная программа повышения квалификации

«Аналитик данных: основы профессии»

72 часа.

**ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ**

**1.Цель программы** – сформировать навык поиска данных, анализа, переработки данных с использованием языка SQL, Python.

**2.Планируемые результаты обучения:**

знать:

архитектуру и структуру базы данных в общем виде;

простые SQL запросы, чтобы получать данные из базы данных (БД);

джоины, чтобы создавать новые таблицы;

аналитические функции SQL, чтобы изучать характеристики данных;

базу данных PostgreSQL, чтобы хранить данные.

уметь:

работать в Python для DataScience и работы с данными;

писать первые самостотятельные программы на языке Python;

контролировать ход работы при помощи условных операторов;

разделять программы на составные части при помощи функций и классов;

получать данные из разных источников с помощью SQL

выстраивать иерархии метрик и проектировать гипотезы

анализировать базы данных и выявлять закономерности между показателями;

интерпретировать данные в понятной для бизнеса форме

составлять интерактивные дашборды

проектировать отчетность под заказчика

обрабатывать данные из разных источников с помощью Power BI или Tableau

владеть:

основами системы контроля версий;

навыком корректно обрабатывать ошибки в программах;

навыком написания простых SQL запросов, чтобы получать данные из базы данных (БД);

импорт и экспорт данных в БД.

**3.Категория слушателей**

3.1. Программа разработана для для слушателей, имеющих высшее или среднее профессиональное образование.

**4.Учебный план программы «**Аналитик данных: основы профессии**»**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль** | **Всего, час** | **Виды учебных занятий** | | |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| **1** | Модуль 1 Работа аналитика в компании | 7 | 7 |  |  |
| **2** | Модуль 2  SQL для аналитиков | 19 | 9 | 10 |  |
| **3** | Модуль 3 Python для аналитиков | 46 | 18 | 28 |  |
|  | Всего |  | 34 | 38 |  |
| **Итоговая аттестация** | |  | **Зачёт** | | |
|  | | **72** |  | | |

**5.Календарный план-график реализации образовательной** программы **«**Аналитик данных: основы профессии**»**

Начало обучения - 01.11.2020

Завершение обучения - 15.11.2020

Периодичность набора - дважды в месяц

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Наименование учебных модулей** | **Трудоёмкость (час)** | **Сроки обучения** |
| 1 | Модуль 1 Работа аналитика в компании | 7 | 01.11.2020-02.11.2020 |
| 2 | Модуль 2  SQL для аналитиков | 19 | 03.11.2020-09.11.2020 |
| 3 | Модуль 3 Python для аналитиков | 46 | 10.11.2020-15.11.2020 |
| Всего: | | 72 |  |

**6.Учебно-тематический план программы «**Аналитик данных: основы профессии**»**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль / Тема** | **Всего, час** | **Виды учебных занятий** | | | **Формы контроля** |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| **1** | **Модуль 1 Работа аналитика в компании** | **7** | **7** |  |  |  |
| 1.1 | Знакомство с бизнес-показателями и понимание бизнеса |  | 1 |  |  |  |
| 1.2 | Финансовые метрики |  | 0,5 |  |  |  |
| 1.3 | Маркетинговые метрики и метрики продукта |  | 1 |  |  |  |
| 1.4 | Иерархия метрик |  | 1 |  |  |  |
| 1.5 | Сбор требований и разработка отчётности |  | 1 |  |  |  |
| 1.6 | Формулирование гипотез |  | 1 |  |  |  |
| 1.7 | Дизайн тестов, проведение и анализ |  | 0,5 |  |  |  |
| 1.8 | Оптимизация отчетности |  | 1 |  |  |  |
| 2 | **Модуль 2**  **SQL для аналитиков** | **29** | **9** | **10** |  |  |
| 2.1 | Введение в SQL.Установка ПО |  | 1 | 1 |  | практическое задание |
| 2.2 | Работа с базами данных |  | 1 | 2 |  | практическое задание |
| 2.3 | Основы SQL |  | 2 | 1 |  | практическое задание |
| 2.4 | Углубление в SQL |  | 2 | 2 |  | практическое задание |
| 2.5 | Работа с PostgresSQL |  | 2 | 2 |  | практическое задание |
| 2.6 | Работа с MongoDB |  | 1 | 2 |  | практическое задание  Итоговое задание по модулю |
| **3** | **Модуль 3 Python для аналитиков** | **46** | **18** | **28** |  |  |
| 3.1 | Основы Python |  | 2 | 2 |  |  |
| 3.2 | Введение в типы и циклы данных |  | 3 | 3 |  | практическое задание |
| 3.3 | Функции |  | 2 | 3 |  | практическое задание |
| 3.4 | Работа с файлами |  | 2 | 3 |  | практическое задание |
| 3.5 | Время в Python |  | 2 | 3 |  | практическое задание |
| 3.6 | Классы в Python |  | 2 | 3 |  | практическое задание |
| 3.7 | NumPy и матрицы |  | 1 | 3 |  | практическое задание |
| 3.8 | Основы Pandas |  | 2 | 3 |  | практическое задание |
| 3.9 | Продвинутый Pandas |  | 1 | 2 |  | практическое задание |
| 3.10 | Объединение датафреймов |  | 1 | 3 |  | итоговое задание по модулю |
|  | Всего | 72 | 34 | 38 |  |  |

**7. Учебная (рабочая) программа повышения квалификации «**Аналитик данных: основы профессии**»**

**Модуль 1 Работа аналитика в компании**

**1.1 Знакомство с бизнес-показателями и понимание бизнеса**

**1.2 Финансовые метрики**

Финансовое планирование, бюджетирование, бизнес-планирование – что это все такое и в чем роль аналитика в этом? Говорим с бухгалтерами и финансистами на одном языке: термины, популярные ошибки, разночтения Кесарю кесарево? Или все таки финансовые метрики должны дополняться продуктовыми / маркетинговыми?

**1.3 Маркетинговые метрики и метрики продукта**

Инструменты сбора и настройки маркетинговой аналитики; Что важно анализировать? Что вводит в заблуждение?; Кейс выявления проблем в показателях (по цифрам все ок, а компания теряет деньги)

**1.4 Иерархия метрик**

Метрики как индикаторы стратегических целей бизнеса. Юнит-экономика; Проектирование и выбор системы метрик (от клиента/продукта); Алерты и системы раннего оповещения; Пирамида метрик.

**1.5 Сбор требований и разработка отчётности**

Определение ответственного/ главного по метрикам; Интервьюирование заказчика / составление пирамиды метрик; Проектирование будущей отчетности / составление структуры; Разработка отчетности.

**1.6 Формулирование гипотез**

Подходы к формулированию гипотез; Что гипотеза, а что нет? А что идея? В чем отличие?

Сколько проверять гипотезу? А если никак не проверяется? Как найти условия, в которых гипотеза будет выполняться?; Всем ли бизнесам подходит проверка гипотез?; Что после проверки надо делать?; Приводят ли гипотезы к росту?; Гипотеза для бизнеса или бизнес для гипотез? Что если текущий бизнес - это лишь площадка для строительства нового?; Фреймворки и кейсы известных компаний.

**1.7 Дизайн тестов, проведение и анализ**

Какие тесты бывают? По типам, целям, инструментам, уровню, исполнителю, динамике

А/Б, А/А, А/С и прочие: когда что применяется и зачем?; Сколько стоит проведение теста и когда его проводить?; Негативные результаты и тесты на ухудшение; Бизнес-модели и фин.модели; Тесты в рамках разных моделей.

**1.8 Оптимизация отчетности**

Когда отчетности слишком много?; Минимально необходимые и полезные отчеты в компании (кейсы).

**Модуль 2 SQL для аналитиков**

**2.1 Введение в SQL.Установка ПО.**

Введение в SQL.Установка ПО; Работа с базами данных; Основы SQL; Углубление в SQL; Работа с PostgresSQL; Работа с MongoDB; SQL-pro

**2.2 Работа с базами данных**

Определение; Устройство; Архитектура и структура данных; Примеры баз данных (БД);

Получение данных; Схемы в БД: понятие нормализации (3NF), примеры схем хранилища (Star Schema, Snowflake Schema, Hybrid Schemas).

**2.3 Основы SQL**

Простые запросы, Join`ы, агрегаты, группировки, вложенные запросы, индексы, фильтрация, агрегация, импорт/экспорт данных; Базовые команды: SELECT, WHERE, JOIN (Inner, Right, Left, Full), ORDER BY, LIKE, IN, BETWEEN, CASE, UNION, INTERSECT, WITH, GROUP BY, HAVING).

**2.4 Углубление в SQL**

Аналитические функции (Sum( ), Count( ), Max( ), Min( ), Avg( ), Rank( ), Dense\_Rank( ), Percent\_Rank( )); Extract, Transform, Load (ETL): Import/Export посредством SQL (.txt файл, .csv файл, в другую БД); Import/Export посредством ETL программ (Talend, Pentaho, Rapidminer, SSIS).

**2.5 Работа с PostgresSQL**

Устройство, принципы работы PostgreSQL; Как пользоваться PostgreSQL; Как загружать и получать данные с этой БД; Разные форматы данных, кроме .txt и .csv, .json.

**2.6 Работа с MongoDB**

Устройство, принципы работы MongoDB; Как данной БД пользоваться; Как загружать и получать данные с этой БД; Разные форматы данных, кроме .txt и .csv, .json.

**Модуль 3** **Python для аналитиков**

**3.1 Основы Python**

Основные принципы работы интерпретатора Python; Дистрибутив Anaconda

Установка пакетов. Возможные проблемы и методы их решения.Обзор jupyter notebook

Основные элементы программирования; Комментарии и markdown; Hello, world; Git и github

**3.2 Введение в типы и циклы данных**

Понятие переменной; Ветвление; Циклы; Список и операция индексирования;

Итерация по коллекции при помощи for; Enumerate для получения индекса; List comprehension; Сортировка и поиск.

**3.3 Функции**

Типы данных; Словарь; Циклы для словарей; Кортеж; dict comprehension; Сортировка и поиск. Функции; Функция как объект; Как работает range в Python 3;

Лямбды (обзорно); Позиционные и именованные аргументы; Значения по умолчанию.

**3.4 Работа с файлами**

Понятие модуля и пакета; Подсистема pip; Краткий обзор стандартной библиотеки;

Установка стороннего модуля; Работа с файлами; Работа с каталогами; Сохранение и восстановление данных; Библиотечный модуль pickle.

**3.5 Время в Python**

Merging & Joining: Concat, Join, pd.merge; работа со временем в Python; время в python: Pandas timestamp, Panas Datetimeindex, Pd.date\_range.

**3.6 Классы в Python**

Класс и экземпляр класса; Данные экземпляра, методы экземпляра и свойства экземпляра;

Понятие атрибута; Понятия базового класса и производного класса; Подмена методов в производном классе; Функция isinstance и ее применение.

**3.7 NumPy и матрицы**

получить представление о многомерных массивах numpy и их преимуществах перед традиционными массивами Python (оптимизация вычислений, множество реализованных операций и тд и тп); узнать про основные функции взаимодействия с numpy массивами; познакомятся с принципами broadcast’а.

**3.8 Основы Pandas**

Понятие массива и его основные характеристики; структура библиотеки; типы данных библиотеки numpy; принципы вычислений;важнейшие стандартные функции.

**3.9 Продвинутый Pandas**

Основные функции pandas; Sort\_values; Sort\_index; Apply; Map

**3.10 Объединение датафреймов**

Добавление и удаление колонок: Работа с данными: поиск нулей (isnull, isna, notnull и тп), Duplicated, Drop\_duplicates, Unique; Операции индексирования в pandas; Объединение условий для индексирования; Группировка данных и метод agg.

**Описание практико-ориентированных заданий и кейсов**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Номер темы/модуля** | **Наименование практического занятия** | **Описание** |
| **7** | 2.1 | Введение в инфраструктуру. Основы баз данных | Развернуть образ, загрузить базу данных и повторить материал занятия на датасете |
| **8** | 2.2 | Работа с базами данных | Самостоятельно прочитать про limit и offset в  postgresql (например тут -  [http://www.postgresqltutorial.com/postgresql-limit/](http://)).  Вывести 10 последних платежей за прокат фильмов  (таблица payment) |
| **9** | 2.3 | Основы SQL | Применение SQL и выполнение простых запросов |
| **10** | 2.4 | Углубление в SQL: аналитические функции. ETL | Спроектировать базу данных для определенных сущностей в соответствии с правилами |
| **11** | 2.5 | Работа с PostgreSQL | Написать запрос, создать материализованное представление с запросом и обновить его |
| **12** | 2.6 | Работа с MongoDB  Итоговое задание по модулю 2 | Создать собственную базу данных  Выполнение контрольного задания по модулю. (п. 8.3) |
|  | 3.1 | Основы Python | Слушателям задается несколько задач для которых они должны написать программы способные выполнить функции, исключить ошибки, написать код и преобразовать данные. |
|  | 3.2 | Введение в типы и циклы данных |
|  | 3.3 | Функции |
|  | 3.4 | Работа с файлами |
|  | 3.5 | Время в Python |
|  | 3.6 | Классы в Python |
|  | 3.7 | NumPy и матрицы |
|  | 3.8 | Основы Pandas |
|  | 3.9 | Продвинутый Pandas |
|  |  |  |

**8.Оценочные материалы по образовательной программе**

**8.1. Вопросы тестирования по модулям**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ модуля** | **Вопросы входного тестирования** | **Вопросы промежуточного тестирования** | **Вопросы итогового тестирования** |
| **1.1** | Что нужно, чтобы зайти в интернет? | **нет** | Что такое реляционные базы данных  Как выглядит запрос, для вывода ВСЕХ значений из таблицы Orders  Какие данные мы получим из этого запроса?  select id, date, customer\_name from Orders  Есть ли ошибка в запросе?  select id, date, customer\_name from Orders where customer\_name = Mike  Что покажет следующий запрос:  select \* from Orders where date between '2017-01-01' and '2017-12-31' |
| **1.2** | Что делает сочетание клавиш CTRL + V? | **нет** |
| **1.3** | Что является признаком отсутствия интернета на устройстве? | **нет** |
| **1.4** | Где используются горячие клавиши? | **нет** |
| **1.5** | Как еще называются горячие клавиши? | **нет** |
| **1.6.** | Чем отличаются поисковики Яндекс и Google? | **нет** |
| **1.7** | Как понять, что перед вами письмо от мошенников? | **нет** |
| **1.8.** | Что такое «Двухфакторная аутентификация»? | **нет** |
| **1.9.** | Чем облачное хранилище данных отличается от хранения данных на компьютере? | **нет** |
| **1.10.** | Как удалить файлы из облачного хранилища? | **нет** |
| **1.11** | Как добавить файл в облачное хранилище? | **нет** |
| **1.12** | Что такое Google Диск? | **нет** |
| **1.13** | Какая из представленных ниже программ является мессенджером? | **нет** |
| **1.14.** | Что такое «демонстрация экрана» в программах для созвонов, например, в Skype? | **нет** |
| **1.15** | В чем разница между программами для созвонов Skype и Zoom? | **нет** |

**8.2.** П**оказатели и критерии оценивания, шкалы оценивания.**

|  |  |  |
| --- | --- | --- |
| Показатели оценки результата | Критерии оценивания | Шкала оценивания |
| Практическое задание | Работа выполнена полностью и сдана в срок  Работа выполнена не полностью/ Работа не сдана | зачтено  не зачтено |
| Итоговое практическое задание по модулю | Работа выполнена полностью и сдана в срок  Работа выполнена не полностью/ Работа не сдана | зачтено  не зачтено |
| Зачет по итогам прохождения программы | Выполнены итоговые практические задания по модулям и сданы в срок  Итоговые практические задания выполнены не по всем модулям/ не выполнены/не сданы в срок | зачтено  не зачтено |

**8.3.**  **Контрольные задания по модулям**

**Модуль 2**

**Текст задания:**

Что необходимо сделать?

1. Подключиться к базе данных **avia**

2. Описание базы данных: <https://edu.postgrespro.ru/bookings.pdf>

3. Работаем в схеме **bookings**

4. Создать запросы, позволяющие ответить на следующие вопросы:

|  |  |
| --- | --- |
| Вопрос | Количество баллов |
| В каких городах больше одного аэропорта? | 10 |
| В каких аэропортах есть рейсы, которые обслуживаются самолетами с максимальной дальностью перелетов? | 15 |
| Были ли брони, по которым не совершались перелеты? | 15 |
| Самолеты каких моделей совершают наибольший % перелетов? | 25 |
| Были ли города, в которые можно добраться бизнес - классом дешевле, чем эконом-классом? | 25 |
| Узнать максимальное время задержки вылетов самолетов | 25 |
| Между какими городами нет прямых рейсов\*? | 35 |
| Между какими городами пассажиры делали пересадки\*? | 35 |
| Вычислите расстояние между аэропортами, связанными прямыми рейсами, сравните с допустимой максимальной дальностью перелетов в самолетах, обслуживающих эти рейсы \*\* | 45 |

\* - пересадка: остановка в аэропорту длительностью менее 1 суток

\*\* - Кратчайшее расстояние между двумя точками A и B на земной поверхности (если принять ее за сферу) определяется зависимостью:

d = arccos {sin(latitude\_a)·sin(latitude\_b) + cos(latitude\_a)·cos(latitude\_b)·cos(longitude\_a - longitude\_b)}, где latitude\_a и latitude\_b — широты, longitude\_a, longitude\_b — долготы данных пунктов, d — расстояние между пунктами, измеряемое в радианах длиной дуги большого круга земного шара.

Расстояние между пунктами, измеряемое в километрах, определяется по формуле:

L = d·R, где R = 6371 км — средний радиус земного шара.

Для расчета расстояния между пунктами, расположенными в разных полушариях (северное-южное, восточное-западное) , знаки (±) у соответствующих параметров (широты или долготы) должны быть разными.

5. Оформить работу

По итогам работы у вас должны быть подготовлены следующие файлы:

* SQL - запросы желательно выгружать в виде \*.sql файлов, либо в текстовом виде
* Описание - \*.pdf, \*.doc и т.д. файл
* ER - диаграмма - в виде скриншота

**Модуль 3**

**Текст задания:** задание к домашней работе можно найти здесь https://github.com/rtriangle/Netology-statistics/blob/master/homework-6.ipynb

Готовое решение необходимо разместить в этой вкладке.**З**

**8.4.**  **Тестирование-не предполагается**

**Практические задания**

Модуль 3

3.1

Задание 1

Даны 2 переменных, в которых хранятся строки произвольной длины: phrase\_1 и phrase\_2.

Напишите код, который проверяет какая из этих строк длиннее.

Задание 2

Дана переменная, в которой хранится четырехзначное число (год). Необходимо написать программу, которая выведет, является ли данный год високосным или обычным.

Задание 3

Необходимо написать программу, которая будет запрашивать у пользователя месяц и дату рождения и выводить соответствующий знак зодиака.

Задание 4

Вам нужно написать программу для подбора упаковок по размерам товара. Размеры хранятся в переменных (в сантиметрах):

3.2

Задание 1

Дана переменная, в которой хранится слово из латинских букв. Напишите код, который выводит на экран:

среднюю букву, если число букв в слове четное;

две средних буквы, если число букв нечетное.

Задание 2

Напишите программу, которая последовательно запрашивает у пользователя числа (по одному за раз) и после первого нуля выводит сумму всех ранее введенных чисел.

адание 3

Мы делаем MVP dating-сервиса, и у нас есть список парней и девушек.

Выдвигаем гипотезу: лучшие рекомендации мы получим, если просто отсортируем имена по алфавиту и познакомим людей с одинаковыми индексами после сортировки! Но мы не будем никого знакомить, если кто-то может остаться без пары

Задание 4

У нас есть список, содержащий информацию о среднедневной температуре в Фаренгейтах за произвольный период по странам (структура данных в примере). Необходимо написать код, который рассчитает среднюю температуру за период в Цельсиях(!) для каждой страны.

3.3

Вам нужно помочь секретарю автоматизировать работу. Для этого нужно написать программу, которая будет на основе хранимых данных исполнять пользовательские команды.

Исходные данные имеют следующую структуру:

1. перечень всех документов

documents = [

{‘type’: ‘passport’, ‘number’: ‘2207 876234’, ‘name’: ‘Василий Гупкин’},

{‘type’: ‘invoice’, ‘number’: ‘11-2’, ‘name’: ‘Геннадий Покемонов’},

{‘type’: ‘insurance’, ‘number’: ‘10006’, ‘name’: ‘Аристарх Павлов’}

]

2.перечень полок, на которых хранятся документы (если документ есть в documents, то он обязательно должен быть и в directories)

3.4

Переведите содержимое файла purchase\_log.txt в словарь purchases вида:

{‘1840e0b9d4’: ‘Продукты’, …}

Для каждого user\_id в файле visit\_log.csv определите третий столбец с категорией покупки (если покупка была, сам файл visit\_log.csv изменять не надо). Запишите в файл funnel.csv визиты из файла visit\_log.csv, в которых были покупки с указанием категории.

Учтите условия на данные:

содержимое purchase\_log.txt помещается в оперативную память компьютера

содержимое visit\_log.csv - нет; используйте только построчную обработку этого файла

3.5

Задание 1

Печатные газеты использовали свой формат дат для каждого выпуска. Для каждой газеты из списка напишите формат указанной даты для перевода в объект datetime:

The Moscow Times - Wednesday, October 2, 2002

The Guardian - Friday, 11.10.13

Daily News - Thursday, 18 August 1977

Задание 2

Дан поток дат в формате YYYY-MM-DD, в которых встречаются некорректные значения:

stream = [‘2018-04-02’, ‘2018-02-29’, ‘2018-19-02’]

Напишите функцию, которая проверяет эти даты на корректность. Т. е. для каждой даты возвращает True (дата корректна) или False (некорректная дата).

3.6

Задание 1

Напишите функцию, которая возвращает название валюты (поле ‘Name’) с максимальным значением курса с помощью сервиса https://www.cbr-xml-daily.ru/daily\_json.js

Задание 2

Добавьте в класс Rate параметр diff (со значениями True или False), который в случае значения True в методах курсов валют (eur, usd итд) будет возвращать не курс валюты, а изменение по сравнению в прошлым значением. Считайте, self.diff будет принимать значение True только при возврате значения курса. При отображении всей информации о валюте он не используется.

Задание 3

Напишите класс Designer, который учитывает количество международных премий. Подсказки в коде занятия в разделе “Домашнее задание задача 3”.

3.7

Задание 1

Напишите функцию, которая принимает на вход строку и проверяет является ли она валидным транспортным номером (1 буква, 3 цифры, 2 буквы, 2-3 цифры). Обратите внимание, что не все буквы кириллического алфавита используются в транспортных номерах.

Задание 2

Напишите функцию, которая будет удалять все последовательные повторы слов из заданной строки при помощи регулярных выражений.

Задание 3

Напишите функцию, которая будет возвращать акроним по переданной в нее строке со словами.

Задание 4

Напишите функцию, которая будет принимать на вход список email-адресов и выводить их распределение по доменным зонам.

3.8

Задание 1

Создайте numpy array с элементами от числа N до 0 (например, для N = 10 это будет array([9, 8, 7, 6, 5, 4, 3, 2, 1, 0])).

Задание 2

Создайте диагональную матрицу с элементами от N до 0. Посчитайте сумму ее значений на диагонали.

Задание 3

Решите систему уравнений:

4x + 2y + z = 4

x + 3y = 12

5y + 4z = -3

3.9

Используйте файл с оценками фильмов ml-latest-small/ratings.csv. Посчитайте среднее время жизни пользователей, которые выставили более 100 оценок. Под временем жизни понимается разница между максимальным и минимальным значением столбца timestamp для данного значения userId.

Задание 2

Дана статистика услуг перевозок клиентов компании по типам (см. файл с кодом занятия). Необходимо сформировать две таблицы:

таблицу с тремя типами выручки для каждого client\_id без указания адреса клиента

аналогичную таблицу по типам выручки с указанием адреса клиента

Обратите внимание, что в процессе объединения таблиц данные не должны теряться.

Задание 3

В задаче сквозной аналитики вам предоставили данные по местоположению пользователей. Т. е. для каждого user\_id известна последовательность координат (широта/долгота), когда они требовались приложению для полноценной работы. Как бы вы добавили эти сведения в таблицу визитов и покупок? Для составления ответа можно использовать вопросы:

У каждого пользователя известен набор координат. А для связывания с визитом или фактом покупки скорее всего потребуется одно-два числа. Как их получить?

Наборы координат одного и того же пользователя могут быть значительно удалены друг от друга. Как это отразится на вопросе расчетах пункта 1?

Какие дополнительные признаки можно получить из координат? Ведь это просто числа, которые сами по себе мало что дают.

**8.5.**  Оценивание результатов обучения происходит в форме зачета. Зачтено - студент выполнил 70% практических заданий по модулю.

**9.Организационно-педагогические условия реализации программы**

**9.1. Кадровое обеспечение программы**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Фамилия, имя, отчество (при наличии)** | **Место основной работы и должность, ученая степень и ученое звание (при наличии)** | **Ссылки на веб-страницы с портфолио (при наличии)** | **Фото в формате jpeg** | **Отметка о полученном согласии на обработку персональных данных** |
| **1** | Николай Хащанов | Fullstack разработчик в Aurora Group |  |  | да |
| **2** | Константин Башевой | Аналитик-разработчик, Яндекс |  |  | да |

**9.2.Учебно-методическое обеспечение и информационное сопровождение**

|  |  |
| --- | --- |
| **Учебно-методические материалы** | |
| Методы, формы и технологии | Методические разработки,  материалы курса, учебная литература |
| Видеолекции для асинхронного обучения | Видеолекции по Модулям  1. Богачев А. Графики, которые убеждают всех / А. Богачев, Издательство АСТ, 290 c. 2. Нафлик К. Данные: визуализируй, расскажи, используй. Сторителлинг в аналитике / К. Нафлик, Манн, Иванов и Фербер, 2020. 288 c.  3. Tufte E.R. Envisioning information / E.R. Tufte, Fourteenth printing-е изд., Cheshire, Connecticut: Graphics Press, 2013. 126 c.  4. Wong D.M. The Wall Street journal guide to information graphics: the dos and don’ts of presenting data, facts, and figures / D.M. Wong, New York: Norton, 2013. 157 c |
| Практические задания |
|
|

|  |  |
| --- | --- |
| **Информационное сопровождение** | |
| Электронные  образовательные ресурсы | Электронные  информационные ресурсы |
|  | Введение в sql. Установка ПО  <https://www.docker.com>  <https://support.microsoft.com/ru-ru/help/15056/windows-32-64-bit-faq>  <https://github.com/docker/kitematic/releases>  <https://dbeaver.io>  <https://postgrespro.ru/docs/postgresql>  <https://github.com/dbeaver/dbeaver/wiki>  Работа с базами данных  <https://www.postgresql.org/docs/9.2/functions-math.html>  <https://www.postgresql.org/docs/9.2/functions-string.html>  <https://www.postgresql.org/docs/9.2/functions-logical.html>  <http://www.postgresqltutorial.com/postgresql-limit/>  <https://www.postgresql.org/docs/9.2/functions.html>  <https://dev.mysql.com/doc/sakila/en/sakila-structure-tables.html>  Основы sql  <https://habr.com/ru/post/268983/>  <http://www.postgresqltutorial.com/postgresql-aggregate-functions/>  <http://www.skillz.ru/dev/php/article-Obyasnenie_SQL_obedinenii_JOIN_INNER_OUTER.html>  Работа с PostgreSQL  <https://habr.com/ru/post/269497/>  <https://medium.com/@hakibenita/be-careful-with-cte-in-postgresql-fca5e24d2119>  <https://postgrespro.ru/docs/postgrespro/9.5/using-explain>  <https://habr.com/ru/post/203320/>  <https://analyticscanvas.com/twb-vs-twbx-in-tableau/>  «Основы работы с Tableau Server». <https://www.tableau.com/pricing/teams-orgs> |
| Компетенции и инструменты аналитика. Роль аналитики в принятии решений  <https://docs.google.com/spreadsheets/d/1_Fre_SjKhqqGquZMOe6AM3hIK3mKB_h16IJDP_FBrFk>  Ошибки и провалы при интерпретации аналитических показателей  [https://ru.wikipedia.org/wiki/Нулевая\_гипотеза](https://ru.wikipedia.org/wiki/)  Разбор кейсов реальных бизнесов. поиск инсайдов в данных  <https://vwo.com/tools/ab-test-siginficance-calculator/>  [https://vwo.com/downloads/ab\_testing\_significance\_calculator.xls](https://vwo.com/tools/ab-test-siginficance-calculator/)  [https://docs.google.com/spreadsheets/d/1U-dwmNhdmuFqmFiJqk8OUZN1OOHA8QeoXS2YlnwZgIQ/edit#gid=702666562](https://vwo.com/tools/ab-test-siginficance-calculator/)  Python для работы с базами данных  [https://www.udacity.com/course/intro-to-relational-databases—ud197](https://www.udacity.com/course/intro-to-relational-databases)  [https://khashtamov.com/ru/postgresql-python-psycopg2/](https://www.udacity.com/course/intro-to-relational-databases)  [https://khashtamov.com/ru/mysql-python/](https://www.udacity.com/course/intro-to-relational-databases)  [https://medium.freecodecamp.org/sqlalchemy-makes-etl-magically-easy-ab2bd0df928](https://www.udacity.com/course/intro-to-relational-databases)  [http://datalytics.ru/all/stanovyas-guru-api-yandeks-metriki/](https://www.udacity.com/course/intro-to-relational-databases)  [https://www.dataquest.io/blog/python-api-tutorial/](https://www.udacity.com/course/intro-to-relational-databases)  [https://www.pythonforbeginners.com/api/list-of-python-apis](https://www.udacity.com/course/intro-to-relational-databases) |

**9.3.Материально-технические условия реализации программы**

|  |  |
| --- | --- |
| Вид занятий | Наименование оборудования,  программного обеспечения |
| Видеолекции | Занятия по программе проводятся с использованием дистанционных образовательных технологий. Каждый слушатель обеспечен доступом к образовательной платформе <https://netology.ru/>. |
| Самостоятельная работа | Jupyter Notebook и Anaconda  <https://medium.com/codingthesmartway-com-blog/getting-started-with-jupyter-notebook-for-python-4e7082bd5d46>  [https://www.dataquest.io/blog/jupyter-notebook-tutorial/](https://medium.com/codingthesmartway-com-blog/getting-started-with-jupyter-notebook-for-python-4e7082bd5d46)  [https://www.analyticsvidhya.com/blog/2018/05/starters-guide-jupyter-notebook/](https://medium.com/codingthesmartway-com-blog/getting-started-with-jupyter-notebook-for-python-4e7082bd5d46)  [https://www.youtube.com/watch?v=YJC6ldI3hWk&feature=youtu.be](https://medium.com/codingthesmartway-com-blog/getting-started-with-jupyter-notebook-for-python-4e7082bd5d46)  Справочная литература  <https://ss64.com/>  [https://stepik.org/lesson/2231/step/1?unit=928](https://ss64.com/)  [https://www.datacamp.com/community/tutorials/data-structures-python](https://ss64.com/) |

**III.Паспорт компетенций (Приложение 2)**

ПАСПОРТ КОМПЕТЕНЦИИ

«Аналитик данных: основы профессии»

Общество с ограниченной ответственностью «Центр онлайн-обучения Нетология-групп»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Наименование компетенции | | Управление информацией и данными | |
| 2. | Указание типа компетенции | общекультурная/  универсальная |  | |
| общепрофессиональная |  | |
| профессиональная | профессиональная | |
| профессионально-специализированная |  | |
| 3. | Определение, содержание и основные сущностные характеристики компетенции | | под компетенцией понимается способность искать, отбирать, систематизировать и передавать информацию с помощью цифровых средств  *знать:*  основы поиска и отбора информации  методы работы с поисковыми алгоритмами  способы обработки данных  форматы сбора данных для их обработки  *уметь:*  использовать поисковые алгоритмы  применять актуальные инструменты для визуализации данных  применять современные методы работы с данными для их структуризации  применять навыки аналитического мышления для систематизации информации  *владеть:*  навыками сбора информации из заданных источников  навыками отбора информации  навыками критического оценивания информации  навыком структуризации и систематизации полученной информации  навыком эффективного использования полученной информации  навыком проверки корректности полученной информации  навыком приоритезации данных в рамках полученной информации | |
| 4. | Дескриптор знаний, умений и навыков по уровням | | Уровни сформированности компетенции  обучающегося | Индикаторы |
|  | | Начальный уровень  (Компетенция недостаточно развита. Частично проявляет навыки, входящие в состав компетенции. Пытается, стремится проявлять нужные навыки, понимает их необходимость, но у него не всегда получается.) | *знать:*  основы поиска и отбора информации  *уметь:*  использовать поисковые алгоритмы  *владеть:*  навыками сбора информации из заданных источников |
|  | | Базовый уровень  (Уверенно владеет навыками, способен, проявлять соответствующие навыки в ситуациях с элементами неопределенности, сложности.) | *знать:*  методы работы с поисковыми алгоритмами  *уметь:*  применять навыки аналитического мышления для систематизации информации  *владеть:*навыками сбора информации из заданных источников  навыками отбора информации |
|  | | Продвинутый  (Владеет сложными навыками, способен активно влиять на происходящее, проявлять соответствующие навыки в ситуациях повышенной сложности.) | *знать:*  способы обработки данных  алгоритмы при работе с полученными из различных источников данными  *уметь:*  применять современные методы работы с данными для их структуризации  *владеть:*  навыками критического оценивания информации  навыком проверки корректности полученной информации |
|  | | Профессиональный  (Владеет сложными навыками, создает новые решения для сложных проблем со многими взаимодействующими факторами, предлагает новые идеи и процессы, способен активно влиять на происходящее, проявлять соответствующие навыки  в ситуациях повышенной сложности.) | *знать:*  способы обработки данных  *уметь:*  применять актуальные инструменты для визуализации данных  применять и оценивать различные табличные процессоры для обработки числовых данных  *владеть:*  навыком приоритезации данных в рамках полученной информации  навыком проверки корректности полученной информации  навыком эффективного использования полученной информации |
| 5. | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции | | Связано с компетенцией Критическое мышление в цифровой среде. Программа предполагает, что слушатель уже владеет данной компетенцией. | |
| 6. | Средства и технологии оценки | | Тестирование | |

**IV.Иная информация о качестве и востребованности образовательной программы** Факультет Нетологии «Аналитика и Data Science» — [обладатель «Премии Рунета»](http://premiaruneta.ru/site/laureats19)в номинации «Образование и кадры» в 2019 году

**V.Рекомендаций к программе от работодателей**: прилагаются

**VI.Указание на возможные сценарии профессиональной траектории граждан** **(Приложение 4)**

**Сценарий профессиональной траектории граждан**

по результатам освоения программы “Визуализация данных без программирования в Tableau”

|  |  |
| --- | --- |
| **Цели получения персонального цифрового сертификата** | |
| **текущий статус** | **цель** |
| **Трудоустройство** | |
| состоящий на учете в Центре занятости | трудоустроенный, самозанятый (фриланс), ИП/бизнесмен |
| безработный |
| безработный по состоянию здоровья |
| **Развитие компетенций в текущей сфере занятости** | |
| работающий по найму в организации, на предприятии | сохранение текущего рабочего места |
| работающий по найму в организации, на предприятии | развитие профессиональных качеств |
| работающий по найму в организации, на предприятии | повышение заработной платы |
| работающий по найму в организации, на предприятии | смена работы без изменения сферы профессиональной деятельности |
| временно отсутствующий на рабочем месте (декрет, отпуск по уходу за ребенком и др.) | сохранение и развитие квалификации |
| **Переход в новую сферу занятости** | |
| освоение новой сферы занятости | самозанятый, ИП/бизнесмен, расширение кругозора |

**VII.Дополнительная информация**