

1.1 Проблема проекта.

На сегодняшний день, доходы жителей России очень сильно разнятся в зависимости от региона. Однако рядовой пользователь может быть не в курсе того, насколько сильна эта дифференциация.

1.2 Актуальность проекта.

На сегодняшний день эта проблема очень актуальна, так как среднедушевые доходы населения – ключевой показатель благосостояния жителей той или иной страны.

Почему была выбрана эта тема? Именно, потому что доходы – ключевой показатель благосостояния жителей России. Эта проблема касается меня и мое окружение напрямую.

В решении проблемы заинтересованы экономисты и географы, а также те люди, которые собираются переезжать на новое место с целью работы. Так же пользователям будет удобно выбрать регион для дальнейшего проживания.

1.3 Цель проекта.

Создание интерактивной карты, которая отображает дифференциацию доходов населения в зависимости от года, за который были проведены подсчеты официальными государственными службами статистики.

1.4 Задачи проекта.

1. Поиск необходимой информации на сайте витрины государственной статистики.
2. Проектирование базы данных на основе .geojson и .csv.
3. Разработка программы, создающей карты в реальном времени, используя стек данных на основе .geojson с помощью Python.
4. Разработка интерактивной карты на основе .geojson с помощью Python и HTML.
5. Тестирование интерактивной карты.
6. Представление проекта аудитории.

В качестве доказательства важности данной темы был проведен небольшой опрос среди 20ти человек. Из них 86% ответили что ознакомлены с проблемой доходов населения в России, среди этих людей только 70 процентов имеют представление о том, насколько сильно доходы неравномерны.

Материал для исследования – статьи экономистов, рассматривающих доходы населения людей разных стран, отражающие отношение ученых и исследователей к проблеме проекта.

Литература подтверждает наличие в России дифференциации и неравного распределения доходов населения по регионам России.

Какими будут последствия решения данной проблемы – у людей появится заинтересованность к теме доходов населения, исследователи, которые занимаются доходами и финансами лучше будут понимать проблему, имея визуальное представление. На глобальном уровне государство будет более замотивировано к решению проблем с доходами населения.

В качестве основного языка программирования для реализации проекта был выбран язык Python 3.11. Python – язык, направленный на Data-аналитику и обладающий множеством библиотек, которые позволяют визуализировать данные, создавать интерактивные решения множества проблем.

Geojson – открытый формат, предназначенный для хранения географических структур данных. С помощью него можно воспроизводить карты и фигуры.

Folium – открытая библиотека для Python, позволяющая создавать интерактивные карты на основе Leaflet.js. С помощью этой библиотеки можно просто и быстро визуализировать данные из различных форматов: geojson, topotiff, geotiff и другие.

Matplotlib – открытая библиотека для Python, позволяющая визуализировать данные с помощью графиков, используя вышеуказанные форматы данных.

Pandas и Geopandas – две схожие библиотеки для Python, позволяющие удобно и легко редактировать, изучать и разрабатывать стеки данных на основе .csv, .json и .geojson файлов.

Выбор пал именно на вышеуказанные технологии из-за их простоты работы с географическими массивами данных. Python поддерживает высокий уровень разработки скриптов для Data-аналитики, на чем и завязан данный проект. Форматы .csv и .geojson очень удобны для работы, так как их редактирование может осуществляться на встроенном софте

csv файл, содержащий стек данных за 11 лет, отображающий доходы населения Российской Федерации в каждом регионе.

.geojson, содержащий стек географических структур России. Он включает себя координаты точек, по которым необходимо строить сложные полилинии, составляющие географический рисунок РФ.

.html файл, содержащий описание интерактивной карты
.jpg файлы, наглядно отражающие дифференциацию доходов населения.

pandas_map.py – Python файл, в котором описан процесс слияния данных из .csv в .geojson и визуализации стека данных за 11 лет на картах.

maps.py – Python файл, в котором реализован скрипт создания интерактивной карты на основе .geojson, полученного в ходе слияния в программе pandas_map.py.

Стек географических данных обрабатывается с помощью библиотеки geopandas, далее он сливается со стеком экономических данных с помощью библиотеки pandas, слияние происходит по соответствию названий регионов.

Затем на основе этих данных библиотекой matplotlib создаются графики, отражающие дифференциацию доходов жителей РФ.

Затем файл, содержащий географический массив информации и экономический массив передается на вход программе maps.py, где в свою очередь, с помощью библиотеки Folium создается html файл, содержащий интерактивную карту со слоями-годами.

Выводы: Разработанная карта очень хорошо отражает уровень доходов населения нашей страны. Она позволяет взаимодействовать с регионами и очень удобно переключаться между слоями.

Графики, которые разработаны как прототип карты – также замечательно отражают уровень доходов, но не обладают интерактивностью, что решается интерактивной версией.

В ходе разработки я освоил навыки работы с такими типами данных как geojson, csv, библиотеками для работы с географическими массивами данных, библиотеками для работы с математическими графиками.

После разработки макетов, я представил прототип проекта своим товарищам, которые оставили положительный опыт использования интерактивной картой