

## **Итоговый проект**

### **Анализ факторов риска, предсказывающих развитие диабета**

#### **Содержание:**

- [Описание проекта](#)
- [Этапы анализа](#)
- [Инструменты анализа](#)
- [Ссылка на данные](#)
- [Описание датасетов](#)
- [Source to target датафрейма diabetes\\_data\\_fin](#)
- [Словарь данных датафрейма diabetes\\_data\\_fin](#)
- [Source to target датафрейма diabetes\\_map\\_2010\\_2015](#)
- [Source to target датафрейма diabetes\\_map\\_2010\\_2015\\_coord](#)
- [Словарь данных датафрейма diabetes\\_map\\_2010\\_2015\\_coord](#)
- [Выводы на основе матрицы корреляции](#)
- [Рекомендации на основе анализа](#)

#### **Описание проекта**

Проект направлен на анализ факторов риска, предсказывающих развитие диабета на основе данных исследования Behavioral Risk Factor Surveillance System (BRFSS) за 2015 год.

Заболевание диабетом имеет серьезные последствия как индивидуально, так и для экономики в целом. В данном проекте акцент будет направлен на анализ факторов риска, связанных с развитием диабета, с использованием обширного набора данных, который включает информацию о медицинских показателях, образе жизни, уровне образования и финансовом состоянии участников.

Проект сосредоточен на выявлении взаимосвязей между различными переменными, такими как уровень холестерина, индекс массы тела, курение, физическая активность и другие аспекты здоровья, которые могут влиять на вероятность возникновения диабета.

Цель проекта понять, как различные параметры образа жизни, демографические характеристики и общее состояние здоровья связаны с риском развития диабета, чтобы предложить более эффективные меры профилактики и рекомендации для улучшения здоровья населения.

#### **Этапы анализа**

1. Выгрузка данных заболевания
2. Преобразование данных заболевания: создание naming conventions, изменение типа данных, дополнение данных
3. Документация (source to target, словарь данных)
4. Корреляционный анализ для оценки силы и направления связей между переменными
5. Выводы на основе матрицы корреляции
6. Сохранение матрицы корреляции

7. Группировка данных заболевания на основе матрицы корреляции
8. Анализ групп и стадий диабета
9. Создание визуализаций для представления взаимосвязей и факторов риска
10. Загрузка данных для карты распространения диабета и доли людей с диабетом, получавших лечение
11. Преобразования данных для карты: выборка данных за период с 2010 по 2015, создание naming conventions, преобразование типа данных, обработка данных
12. Загрузка данных координат стран
13. Преобразование данных координат: выборка полигонов, переименование части стран
14. Объединение датасета для карты и датасета с координатами
15. Удаление пустых значений из датасета для карты
16. Выгрузка датасета в csv
17. Документация (source to target, словарь данных)
18. Создание визуализаций на основе данных для карты
19. Рекомендации на основе анализа

### **Инструменты анализа**

1. Visual studio code
  - a. библиотеки: pandas, os, kagglehub, numpy, matplotlib.pyplot, seaborn
2. DataLens

### **Ссылка на данные**

<https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset/data>  
<https://www.ncdrisc.org/data-downloads-diabetes.html>  
<https://yandex.cloud/ru/docs/datalens/dataset/data-types>

### **Описание датасетов**

Исследование Behavioral Risk Factor Surveillance System (BRFSS) за 2015 год, включает в себя 253 680 ответов участников и 22 переменные с данными о здоровье и факторах риска, в данных есть проблема дисбаланса классов поэтому они не подходят для машинного обучения.

Файл NCD RisC Lancet 2024 Diabetes age specific countries содержит данные по странам, стандартизированные по возрасту, о приросте диабета у людей от 18 лет и старше и о доле людей с диабетом, которые получали лечение от 30 лет и старше.

Файл all country borders содержит данные о координатах стран.

Naming convention:

- латинские буквы
- нижний регистр
- разделитель «\_»

### Source to target датафрейма diabetes\_data\_fin

- данные из diabetes\_012\_health\_indicators\_BRFSS2015.csv были загружены в таблицу diabetes\_data\_raw

Source				Target			
Название таблицы	Атрибут	Тип данных	Логика преобразования	Название таблицы	Атрибут	Тип данных	Пример данных
diabetes_data_raw	Diabetes_012	float64	-	diabetes_data_fin	diabetes_012	float64	0.0
diabetes_data_raw	HighBP	float64	-	diabetes_data_fin	high_blood_pressure	float64	1.0
diabetes_data_raw	HighChol	float64	-	diabetes_data_fin	high_cholesterol	float64	1.0
diabetes_data_raw	CholCheck	float64	-	diabetes_data_fin	cholesterol_check	float64	1.0
diabetes_data_raw	BMI	float64	-	diabetes_data_fin	body_mass_index	float64	40.0
diabetes_data_raw	Smoker	float64	-	diabetes_data_fin	smoker	float64	1.0
diabetes_data_raw	Stroke	float64	-	diabetes_data_fin	stroke	float64	0.0
diabetes_data_raw	HeartDisease orAttack	float64	-	diabetes_data_fin	heart_disease_or_attack	float64	0.0
diabetes_data_raw	PhysActivity	float64	-	diabetes_data_fin	physical_activity	float64	0.0
diabetes_data_raw	Fruits	float64	-	diabetes_data_fin	fruits	float64	0.0
diabetes_data_raw	Veggies	float64	-	diabetes_data_fin	veggies	float64	0.0
diabetes_data_raw	HvyAlcohol Consump	float64	-	diabetes_data_fin	heavy_alcohol_consump	float64	0.0

diabetes_data_raw	AnyHealthcare	float64	-	diabetes_data_fin	any_health_care	float64	1.0
diabetes_data_raw	NoDocbcCost	float64	-	diabetes_data_fin	no_doc_bc_cost	float64	0.0
diabetes_data_raw	GenHlth	float64	-	diabetes_data_fin	general_health	float64	5.0
diabetes_data_raw	MentHlth	float64	-	diabetes_data_fin	mental_health	float64	18.0
diabetes_data_raw	PhysHlth	float64	-	diabetes_data_fin	physical_health	float64	15.0
diabetes_data_raw	DiffWalk	float64	-	diabetes_data_fin	diff_walking	float64	1.0
diabetes_data_raw	Sex	float64	-	diabetes_data_fin	sex	float64	0.0
diabetes_data_raw	Age	float64	-	diabetes_data_fin	age	float64	9.0
diabetes_data_raw	Education	float64	-	diabetes_data_fin	education	float64	4.0
diabetes_data_raw	Income	float64	-	diabetes_data_fin	income_usd	float64	3.0
-	-	-	diabetes_data_raw['diabetes_desc'] = diabetes_data_raw['diabetes_012'].map({0: 'нет', 1: 'предиабет', 2: 'диабет'})	diabetes_data_fin	diabetes_desc	object	нет
-	-	-	diabetes_data_raw['high_blood_pressure_desc'] = diabetes_data_raw['high_blood_pressure'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	high_blood_pressure_desc	object	да
-	-	-	diabetes_data_raw['high_cholesterol_desc'] = diabetes_data_raw['high_cholesterol'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	high_cholesterol_desc	object	да

-	-	-	diabetes_data_raw['cholesterol_check_desc'] = diabetes_data_raw['cholesterol_check'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	cholesterol_check_desc	object	да
-	-	-	diabetes_data_raw['smoker_desc'] = diabetes_data_raw['smoker'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	smoker_desc	object	да
-	-	-	diabetes_data_raw['stroke_desc'] = diabetes_data_raw['stroke'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	stroke_desc	object	нет
-	-	-	diabetes_data_raw['heart_disease_or_attack_desc'] = diabetes_data_raw['heart_disease_or_attack'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	heart_disease_or_attack_desc	object	нет
-	-	-	diabetes_data_raw['physical_activity_desc'] = diabetes_data_raw['physical_activity'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	physical_activity_desc	object	нет
-	-	-	diabetes_data_raw['fruits_desc'] = diabetes_data_raw['fruits'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	fruits_desc	object	нет
-	-	-	diabetes_data_raw['veggies_desc'] = diabetes_data_raw['veggies'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	veggies_desc	object	да
-	-	-	diabetes_data_raw['heavy_alcohol_consump_desc'] = diabetes_data_raw['heavy_alcohol_consump'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	heavy_alcohol_consump_desc	object	нет
-	-	-	diabetes_data_raw['any_health_care_desc'] =	diabetes_data_fin	any_health_care_desc	object	да

			diabetes_data_raw['any_health_care'].map({0: 'нет', 1: 'да'})				
-	-	-	diabetes_data_raw['no_doc_bc_cost_desc'] = diabetes_data_raw['no_doc_bc_cost'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	no_doc_bc_cost_desc	object	нет
-	-	-	diabetes_data_raw['general_health_desc'] = diabetes_data_raw['general_health'].map({1: 'отлично', 2: 'очень хорошо', 3: 'хорошо', 4: 'нормально', 5: 'плохо'})	diabetes_data_fin	general_health_desc	object	плохо
-	-	-	diabetes_data_raw['diff_walking_desc'] = diabetes_data_raw['diff_walking'].map({0: 'нет', 1: 'да'})	diabetes_data_fin	diff_walking_desc	object	да
-	-	-	diabetes_data_raw['sex_desc'] = diabetes_data_raw['sex'].map({0: 'женщины', 1: 'мужчины'})	diabetes_data_fin	sex_desc	object	женщины
-	-	-	diabetes_data_raw['age_desc'] = diabetes_data_raw['age'].map({1: 'от 18 до 24', 2: 'от 25 до 29', 3: 'от 30 до 34', 4: 'от 35 до 39', 5: 'от 40 до 44', 6: 'от 45 до 49', 7: 'от 50 до 54', 8: 'от 55 до 59', 9: 'от 60 до 64', 10: 'от 65 до 69', 11: 'от 70 до 74', 12: 'от 75 до 79', 13: 'от 80 или старше'})	diabetes_data_fin	age_desc	object	от 60 до 64
-	-	-	diabetes_data_raw['education_desc'] = diabetes_data_raw['education'].map({1: 'никогда не посещал школу или	diabetes_data_fin	education_desc	object	12 классов (выпускник

			только детский сад', 2: 'классы с 1 по 8 (начальный)', 3: 'классы с 9 по 11 (средняя школа)', 4: '12 классов (выпускник средней школы)', 5: 'колледж от 1 года до 3 лет (колледж или техническое училище) ', 6: 'колледж 4 года или более (выпускник колледжа)'))				средней школы)
-	-	-	diabetes_data_raw['income_usd_desc'] = diabetes_data_raw['income_usd'].map({1: '< 10 000', 2: '< 15 000', 3: '<20 000', 4: '< 25 000', 5: '< 35 000', 6: '< 50 000', 7: '< 75 000', 8: '75 000 >', })	diabetes_data_fin	income_usd_desc	object	<20 000

### Словарь данных датафрейма diabetes\_data\_fin

Название таблицы	Название атрибута	Тип данных	Бизнес-определение	Пример данных	Уровень доступа
diabetes_data_fin	diabetes_012	float64	стадия диабета 0 – нет, 1 – предиабет, 2 – диабет	0.0	неограниченный
diabetes_data_fin	high_blood_pressure	float64	есть ли высокое давление 0 – нет, 1 – да	1.0	неограниченный
diabetes_data_fin	high_cholesterol	float64	есть ли высокий холестерин 0 – нет, 1 – да	1.0	неограниченный
diabetes_data_fin	cholesterol_check	float64	проверялся ли холестерин в течение последних 5 лет 0 – нет, 1 – да	1.0	неограниченный
diabetes_data_fin	body_mass_index	float64	индекс массы тела	40.0	неограниченный

diabetes_data_fin	smoker	float64	выкуривали хотя бы 100 сигарет на протяжении всей жизни (100 сигарет = 5 пачек сигарет) 0 – нет, 1 – да	1.0	неограниченный
diabetes_data_fin	stroke	float64	говорили ли когда либо, что был инсульт 0 – нет, 1 – да	0.0	неограниченный
diabetes_data_fin	heart_disease_or_attack	float64	есть ли ишемическая болезнь сердца или инфаркт миокарда 0 – нет, 1 – да	0.0	неограниченный
diabetes_data_fin	physical_activity	float64	занимались ли физической активностью последние 30 дней (исключая работу) 0 – нет, 1 – да	0.0	неограниченный
diabetes_data_fin	fruits	float64	употребляете ли один или более фруктов в день 0 – нет, 1 – да	0.0	неограниченный
diabetes_data_fin	veggies	float64	употребляете ли один или более овощей в день 0 – нет, 1 – да	0.0	неограниченный
diabetes_data_fin	heavy_alcohol_consump	float64	сильная алкогольная зависимость: - взрослые мужчины: более 14 напитков в неделю - взрослые женщины: более 7 напитков в неделю 0 – нет, 1 – да	0.0	неограниченный
diabetes_data_fin	any_health_care	float64	имеется ли медицинская страховка (любой вид) 0 – нет, 1 – да	1.0	неограниченный
diabetes_data_fin	no_doc_bc_cost	float64	были ли за последние 12 месяцев случаи когда нужно было обратиться к врачу, но не обращались из-за высокой стоимости 0 – нет, 1 – да	0.0	неограниченный
diabetes_data_fin	general_health	float64	оценка уровня здоровья (шкала от 1-5): 1 – отлично 2 – очень хорошо 3 – хорошо 4 – нормально 5 – плохо	5.0	неограниченный



diabetes_data_fin	mental_health	float64	оценка психического здоровья, стрессы, депрессии, проблемы с эмоциями, сколько дней в течение последних 30 дней психическое здоровье было не очень хорошим? (шкала в днях 1-30)	18.0	неограниченный
diabetes_data_fin	physical_health	float64	оценка физического здоровья: травмы и заболевания, сколько дней в течение последних 30 дней физическое здоровье было не очень хорошим? (шкала в днях 1-30)	15.0	неограниченный
diabetes_data_fin	diff_walking	float64	есть серьезные проблемы с ходьбой или подъемом по ступеням 0 – нет, 1 – да	1.0	неограниченный
diabetes_data_fin	sex	float64	пол 0 – женщина, 2 – мужчина	0.0	неограниченный
diabetes_data_fin	age	float64	возрастные категории:  1 – от 18 до 24 2 – от 25 до 29 3 – от 30 до 34 4 – от 35 до 39 5 – от 40 до 44 6 – от 45 до 49 7 – от 50 до 54 8 – от 55 до 59 9 – от 60 до 64 10 – от 65 до 69 11 – от 70 до 74 12 – от 75 до 79 13 – от 80 или старше	9.0	неограниченный
diabetes_data_fin	education	float64	уровень образования (шкала 1-6):	4.0	неограниченный

			1 - никогда не посещал школу / только детский сад 2 – классы с 1 по 8 (начальный) 3 – классы с 9 по 11 (средняя школа) 4 – 12 класс или GED (выпускник средней школы) 5 – колледж от 1 года до 3 лет (колледж или техническое училище) 6 – колледж 4 года или более (выпускник колледжа)		
diabetes_data_fin	income_usd	float64	шкала доходов (шкала 1-8):  1 – менее 10 000 долларов США 2 – менее 15 000 долларов США 3 – менее 20 000 долларов США 4 – менее 25 000 долларов США 5 – менее 35 000 долларов США 6 – менее 50 000 долларов США 7 – менее 75 000 долларов США 8 – 75 000 долларов США или более	3.0	неограниченный
diabetes_data_fin	diabetes_012	object	стадия диабета	no_diabetes	неограниченный
diabetes_data_fin	high blood pressure	object	есть ли высокое давление	yes	неограниченный
diabetes_data_fin	high_cholesterol	object	есть ли высокий холестерин	yes	неограниченный
diabetes_data_fin	cholesterol_check	object	проверялся ли холестерин в течение последних 5 лет	yes	неограниченный
diabetes_data_fin	body mass index	int64	индекс массы тела	40	неограниченный
diabetes_data_fin	smoker	object	выкуривали хотя бы 100 сигарет на протяжении всей жизни (100 сигарет = 5 пачек сигарет)	yes	неограниченный
diabetes_data_fin	stroke	object	говорили ли когда либо, что был инсульт	no	неограниченный

diabetes_data_fin	heart_disease_or_attack	object	есть ли ишемическая болезнь сердца или инфаркт миокарда	no	неограниченный
diabetes_data_fin	physical_activity	object	занимались ли физической активностью последние 30 дней (исключая работу)	no	неограниченный
diabetes_data_fin	fruits	object	употребляете ли один или более фруктов в день	no	неограниченный
diabetes_data_fin	veggies	object	употребляете ли один или более овощей в день	no	неограниченный
diabetes_data_fin	heavy_alcohol_consump	object	сильная алкогольная зависимость: - взрослые мужчины: более 14 напитков в неделю - взрослые женщины: более 7 напитков в неделю	no	неограниченный
diabetes_data_fin	any_health_care	object	имеется ли медицинская страховка (любой вид)	yes	неограниченный
diabetes_data_fin	no_doc_bc_cost	object	были ли за последние 12 месяцев случаи когда нужно было обратиться к врачу, но не обращались из-за высокой стоимости	no	неограниченный
diabetes_data_fin	general_health	object	оценка уровня здоровья (шкала от 1-5): 1 – отлично 2 – очень хорошо 3 – хорошо 4 – нормально 5 – плохо	poor	неограниченный
diabetes_data_fin	mental_health	int64	оценка психического здоровья, стрессы, депрессии, проблемы с эмоциями, сколько дней в течение последних 30 дней психическое здоровье было не очень хорошим?	18	неограниченный

			(шкала в днях 1-30)		
diabetes_data_fin	physical_health	int64	оценка физического здоровья: травмы и заболевания, сколько дней в течение последних 30 дней физическое здоровье было не очень хорошим? (шкала в днях 1-30)	15	неограниченный
diabetes_data_fin	diff_walking	object	есть серьезные проблемы с ходьбой или подъемом по ступеням	yes	неограниченный
diabetes_data_fin	sex	object	пол	female	неограниченный
diabetes_data_fin	age	object	возрастные категории:  1 – от 18 до 24 2 – от 25 до 29 3 – от 30 до 34 4 – от 35 до 39 5 – от 40 до 44 6 – от 45 до 49 7 – от 50 до 54 8 – от 55 до 59 9 – от 60 до 64 10 – от 65 до 69 11 – от 70 до 74 12 – от 75 до 79 13 – от 80 или старше	60-64	неограниченный
diabetes_data_fin	education	object	уровень образования (шкала 1-6):  1 - никогда не посещал школу / только детский сад 2 – классы с 1 по 8 (начальный) 3 – классы с 9 по 11 (средняя школа) 4 – 12 класс или GED (выпускник средней школы)	high_school_graduate	неограниченный

			5 – колледж от 1 года до 3 лет (колледж или техническое училище) 6 – колледж 4 года или более (выпускник колледжа)		
diabetes_data_fin	income_usd	object	шкала доходов (шкала 1-8):  1 – менее 10 000 долларов США 2 – менее 15 000 долларов США 3 – менее 20 000 долларов США 4 – менее 25 000 долларов США 5 – менее 35 000 долларов США 6 – менее 50 000 долларов США 7 – менее 75 000 долларов США 8 – 75 000 долларов США или более	<20_000	неограниченный

**Source to target датафрем diabetes\_map 2010 2015**

- данные из NCD\_RisC\_Lancet\_2024\_Diabetes\_age\_standardised\_countries.csv были загружены в таблицу diabetes\_map\_2010\_2015
- выбраны данные за период с 2010 по 2015 год

Source				Target			
Название таблицы	Атрибут	Тип данных	Логика преобразования	Название таблицы	Атрибут	Тип данных	Пример данных
map_age_standard	Country/Region/World	object	diabetes_map_2010_2015.country = diabetes_map_2010_2015.country.replace({ "Cote d'Ivoire": 'Ivory Coast', 'Guinea Bissau': 'Guinea-Bissau', 'Lao PDR': 'Laos' })	diabetes_map_2010_2015	country	object	Afghanistan
map_age_standard	ISO	object	-	diabetes_map_2010_2015	sex	object	Men

map_age_standard	Sex	object	-	diabetes_map_2010_2015	year	datetime64	2010-01-01
map_age_standard	Year	int64	map_age_standard['year'] = pd.to_datetime(map_age_standard['year'].astype(str) + '-01-01')	diabetes_map_2010_2015	prevalence_of_diabetes_age>18	float64	0.162883
map_age_standard	Prevalence of diabetes (18+ years)	float64	-	diabetes_map_2010_2015	proportion_of_people_with_diabetes_who_were_treated_age>30	float64	0.231176
map_age_standard	Proportion of people with diabetes who were treated (30+ years)	float64	-	-	-	-	-

#### [Source to target датафрeм diabetes\\_map\\_2010\\_2015\\_coords](#)

- данные из geo\_for\_datalens2/all\_country\_borders.csv были загружены в таблицу geopoligon
- выбраны данные all\_country\_borders\_poly
- изменены названия стран
  - geopoligon.name\_en = geopoligon.name\_en.replace({
 'Russia': 'Russian Federation',
 'Democratic Republic of the Congo': 'DR Congo',
 'Congo-Brazzaville': 'Congo',
 'Côte d'Ivoire': 'Ivory Coast',
 'Turkey': 'Turkiye',
 'United States': 'United States of America',
 'Vietnam': 'Viet Nam',
 'The Bahamas': 'Bahamas',
 'Brunei': 'Brunei Darussalam',
 'Cape Verde': 'Cabo Verde',
 'The Gambia': 'Gambia',
 })

'São Tomé and Príncipe': 'Sao Tome and Principe',  
'Syria': 'Syrian Arab Republic',  
'East Timor': 'Timor-Leste'},})

- объединены данные датафрейма diabetes\_map\_2010\_2015 и датафрейма geopoligon
  - diabetes\_map\_2010\_2015\_coord = pd.merge(diabetes\_map\_2010\_2015, geopoligon, left\_on='country', right\_on='name\_en', how='left')

Source				Target			
Название таблицы	Атрибут	Тип данных	Логика преобразования	Название таблицы	Атрибут	Тип данных	Пример данных
diabetes_map_2010_2015	country	object	-	diabetes_map_2010_2015_coords	country	object	Afghanistan
diabetes_map_2010_2015	sex	object	-	diabetes_map_2010_2015_coords	sex	object	Men
diabetes_map_2010_2015	year	datetime64	-	diabetes_map_2010_2015_coords	year	datetime64	2010-01-01
diabetes_map_2010_2015	prevalence_of_diabetes_age>18	float64	-	diabetes_map_2010_2015_coords	prevalence_of_diabetes_age>18	float64	0.162883
diabetes_map_2010_2015	proportion_of_people_with_diabetes_who_were_treated_age>30	float64	-	diabetes_map_2010_2015_coords	proportion_of_people_with_diabetes_who_were_treated_age>30	float64	0.231176
diabetes_map_2010_2015	name_ru	object	-	diabetes_map_2010_2015_coords	name_ru	object	Афганистан
diabetes_map_2010_2015	name_en	object	diabetes_map_2010_2015_coord = diabetes_map_2010_2015_coord[~(diabetes_map_2010_2015_coord['coords'].isnull())	diabetes_map_2010_2015_coords	coords	object	[[[36.925945, 73.841253], [36.89082

			(diabetes_map_2010_2015_coord['coords'] == "")]				4,73.815059],...
diabetes_map_2010_2015	coords_type	object	-	-	-	-	-
diabetes_map_2010_2015	cords	object	-	-	-	-	-

#### Словарь данных датафрейма diabetes\_map\_2010\_2015\_coord

Название таблицы	Название атрибута	Тип данных	Бизнес-определение	Пример данных	Уровень доступа
diabetes_map_2010_2015_coords	country	object	название страны на английском	Afghanistan	неограниченный
diabetes_map_2010_2015_coords	sex	object	пол	Men	неограниченный
diabetes_map_2010_2015_coords	year	datetime64	год	2010-01-01	неограниченный
diabetes_map_2010_2015_coords	prevalence_of_diabetes_age>18	float64	прирост заболевших диабетом, возраст больше 18	0.162883	неограниченный
diabetes_map_2010_2015_coords	proportion_of_people_with_diabetes_who_were_treated_age>30	float64	доля людей с диабетом, которые получали лечение, возраст больше 30	0.231176	неограниченный
diabetes_map_2010_2015_coords	name_ru	object	название страны на русском	Афганистан	неограниченный
diabetes_map_2010_2015_coords	cords	object	координаты (полигон)	[[[36.925945, 73.841253],[36.890824, 73.815059],...	неограниченный



## Выводы на основе матрицы корреляции

Матрица подтверждает типичную клинико-эпидемиологическую картину:

- диабет тесно связан с возрастом, избыточным весом, сердечно-сосудистыми заболеваниями и ухудшением общего физического состояния
- факторы здорового образа жизни и более высокий социально-экономический статус проявляют защитный эффект
- эти результаты наглядно иллюстрируют, что в программах профилактики и контроля диабета следует особое внимание уделять снижению массы тела, повышению физической активности, улучшению качества питания и решению социально-экономических барьеров в доступе к качественной медпомощи.

На основании матрицы можно выделить несколько самых сильных связей с наличием диабета (переменная `diabetes_012`), а также характер взаимосвязей между другими показателями:

1. самая высокая прямая корреляция у диабета наблюдается с самооценкой общего здоровья (`general_health`) и с наличием серьёзных проблем при ходьбе или подъёме по ступеням (`diff_walking`)
  - поскольку в шкале `general_health` «1» означает «отлично», а «5» – «плохо», положительная корреляция (около  $+0.35...+0.40$ ) говорит о том, что у людей с диагностированным диабетом чаще встречаются худшие оценки общего состояния здоровья
  - похожим образом высокая корреляция с `diff_walking` (примерно  $+0.30$ ) отражает, что ограниченная подвижность существенно связана с диабетом, вероятно, и как причина (малоактивный образ жизни), и как следствие (сосудистые осложнения)
2. выраженные прямые связи обнаруживаются между диабетом и классическими сопутствующими факторами риска, корреляции этих признаков с диабетом лежат в диапазоне примерно от  $+0.20$  до  $+0.30$  и отражают хорошо известное сочетание метаболических и сердечно-сосудистых рисков:
  - возраст (`age`),
  - индексом массы тела (`body_mass_index`),
  - гипертонией (`high_blood_pressure`),
  - ишемической болезнью сердца или перенесённым инфарктом (`heart_disease_or_attack`),
  - высоким уровнем холестерина (`high_cholesterol`) и перенесённым инсультом (`stroke`)
3. умеренно отрицательная связь диабета с уровнем физической активности (`physical_activity`) и с регулярным потреблением овощей (`veggies`)
  - отрицательная корреляция порядка  $-0.10...-0.15$  говорит о том, что более активный образ жизни и здоровое питание ассоциируются с меньшей вероятностью выявления диабета
  - аналогичный, но чуть более слабый эффект даёт потребление фруктов (`fruits`)

4. среди социально-демографических факторов наиболее выраженные обратные связи с диабетом прослеживаются у уровня образования (education) и у дохода (income\_usd)
  - чем выше образование и доход, тем ниже распространённость диабета в выборке
  - более низкий социально-экономический статус сопровождается большим риском хронической болезни
5. переменные, такие как курение (smoker), проверка холестерина (cholesterol\_check), наличие страховки (any\_health\_care) или тяжёлая алкогольная зависимость (heavy\_alcohol\_consump), либо демонстрируют близкие к нулю корреляции, либо слишком разнородны для однозначных выводов без дополнительного углублённого анализа.

### Рекомендации на основе анализа

1. Таргетированный скрининг и раннее выявление
  - a. На основе корреляционного анализа внедрить программы прицельного обследования для людей старше 30 лет и с ИМТ > 25
  - b. На основе корреляционного анализа внедрить программы прицельного обследования для людей старше 30 лет и с ИМТ > 25.
2. Пропаганда сбалансированного питания
  - a. Сократить потребление рафинированных углеводов и трансжиров, увеличить долю клетчатки, овощей и нежирного белка
  - b. Организовать бесплатные мастер-классы по планированию здорового рациона
3. Развитие программ регулярной физической активности
  - a. Стимулировать ежедневные 30–45-минутные аэробные и силовые тренировки (ходьба, велосипед, групповые занятия)
  - b. Интегрировать фитнес в школьные и корпоративные расписания
4. Образовательные кампании и поддержка самоконтроля
  - a. Разрабатывать мультимедийные курсы и мобильные приложения с учётом демографии и языковых особенностей целевых групп
  - b. Обучать методам ведения дневника питания и мотивационным техникам