

МИНОБРНАУКИ РОССИИ
РГУ НЕФТИ И ГАЗА (НИУ) ИМЕНИ И.М. ГУБКИНА
ФАКУЛЬТЕТ АВТОМАТИКИ И ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ
КАФЕДРА ИНФОРМАТИКИ
ДИСЦИПЛИНА «ОСНОВЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ И МАШИН-
НОЕ ОБУЧЕНИЕ»

О Т Ч Е Т
по лабораторной работе №2
«Инструменты анализа данных в отчетах Power BI Desktop»

Выполнила: студентка группы АА-19-05

Данилова М.А.

Проверила: доцент Вишневская Е. А.

Москва 2022

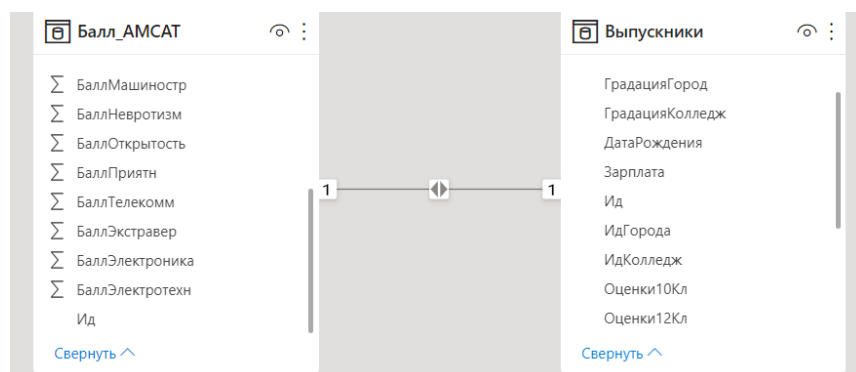
Пошаговая демонстрация выполнения заданий №№6-10 с пояснениями.

6. Для подготовленного набора данных создать отчет с использованием различных элементов визуализаций.

Для анализа были выбраны данные с сайта Kaggle - системы организации конкурсов по исследованию данных (сети специалистов по обработке данных и машинному обучению). Набор данных - Engineering Graduate Salary Prediction (прогноз заработной платы индийских выпускников инженерных специальностей): <https://www.kaggle.com/datasets/manishkc06/engineering-graduate-salary-prediction>

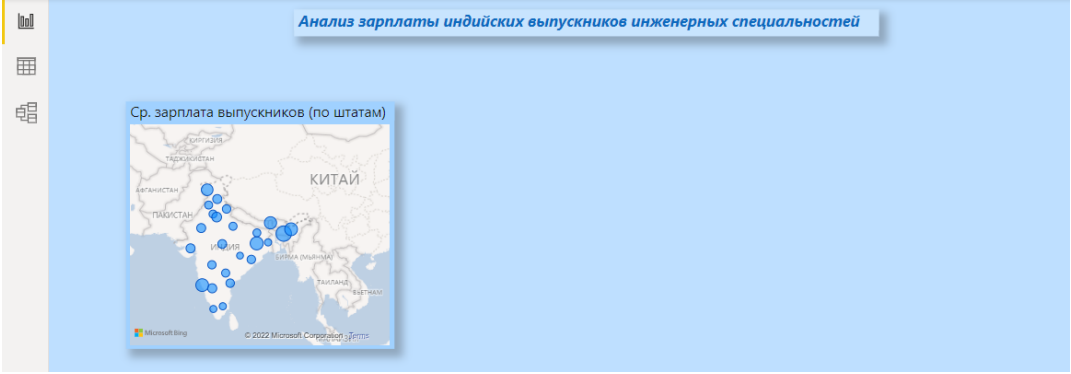
Данные представляют собой таблицы Выпускники и Балл_АМСАТ. Таблица Выпускники включает такие столбцы, как ИД, пол, дата рождения, специализация, ученая степень, зарплата, год выпуска (бакалавриат), штат, оценки за экзамены в 10 классе (общие), идентификатор колледжа/университета, совокупный средний балл по окончанию колледжа и тд. Балл_АМСАТ включает баллы выпускника по различным секциям: в английской, логической, в разделе машиностроения, в секции электротехники, компьютерного программирования, гражданского строительства, личностные тесты и тд. В общей сложности 34 колонки.

Экспортируем данные из Excel файла. Настроим связи между таблицами по идентификатору выпускника:

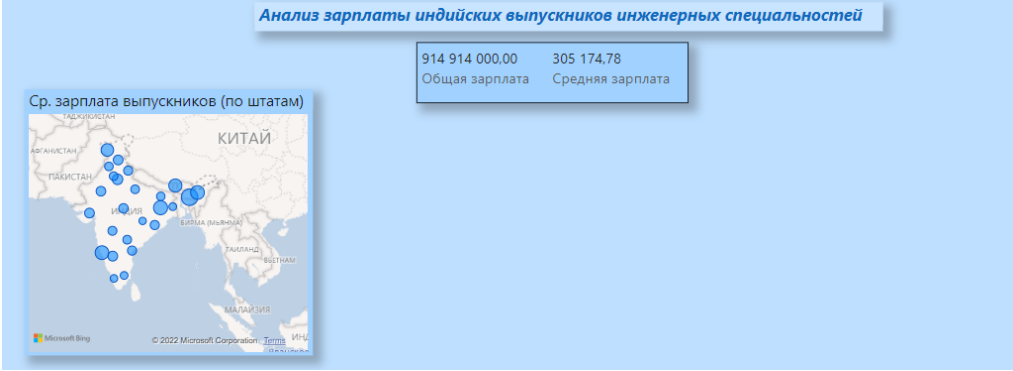


Далее на вкладке «Отчет» добавим визуальные элементы для данных. Изучим зависимость средней заработной платы от штата, года выпуска из бакалавриата, пола, ученой степени и специализации.

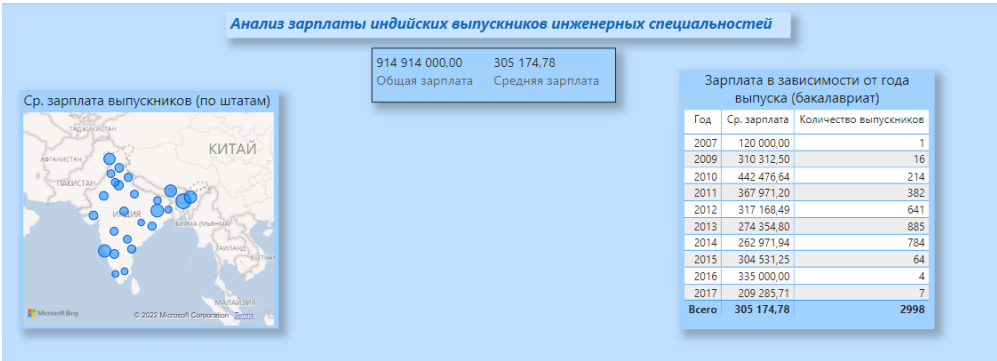
Добавим Карту для отображения штатов, в которых средняя зарплата выпускников является наибольшей (поле Расположение=Таблица Выпускники поле Штат; поле Размер=Таблица выпускники Зарплата-среднее значение), а также текстовое поле с названием. Отформатируем элементы (зададим цвет фона, тень, подписи).



Видно, что средняя зарплата по штатам имеет небольшие колебания. Добавим карточку с данными по суммарной и средней зарплате всех выпускников (применим форматирование по образцу из вкладки Главная).

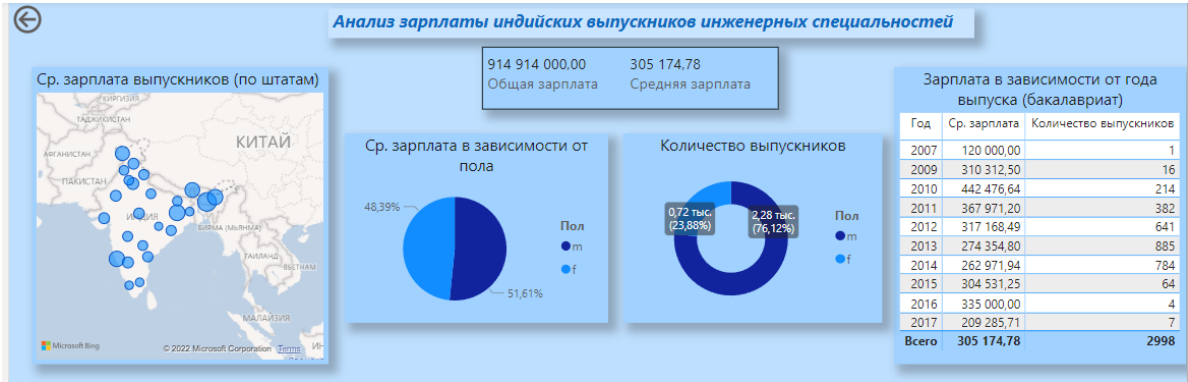


Также отобразим среднюю зарплату выпускников в зависимости от года выпуска из бакалавриата и общее количество выпускников в определенный год (поля элемента Таблица заполним полями ГодвыпускаБак, Зарплата-среднее, Зарплата – количество)



Больше всего выпускников было в 2013 году, однако наибольшая средняя зарплата была в 2010.

Добавим круговую диаграмму и кольцевой график для отображения количества выпускников мужского, женского пола и зависимости средней зарплаты от пола (условные обозначения = поле Пол; также применим форматирование по образцу).



Видно, что мужчин-выпускников было больше, но на среднюю зарплату пол практически не влияет.

Для изучения зависимости средней зарплаты от ученой степени и специализации добавим линейчатую диаграмму с группировкой и гистограмму с группировкой (заполнив соответственно поле ось = ученой степени и поле ось = специализация).

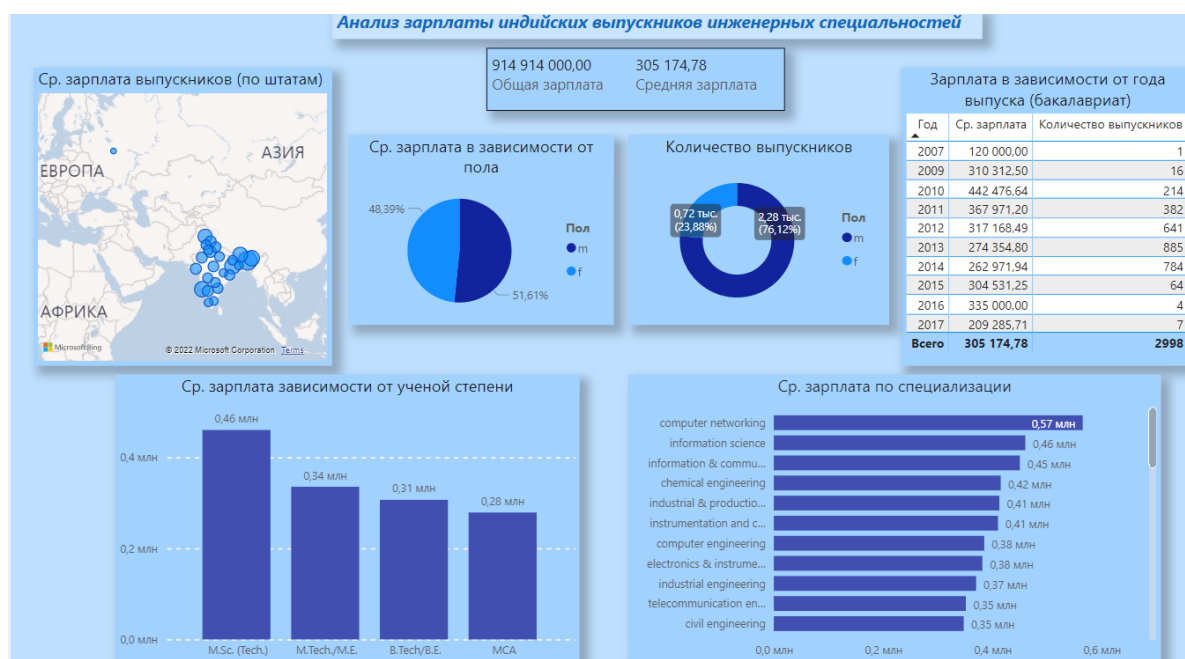


7. Провести анализ данных с использованием сортировок, условного форматирования в таблицах и диаграммах, добавления срезов и фильтров. Применить для анализа инструменты панели «Аналитика».

Отсортируем данные в линейчатой диаграмме и гистограмме по величине средней зарплаты. Для этого в дополнительных параметрах выбранного элемента выберем пункт «Сортировка группы ось» - «Ср.зарплата» - «Сортировка по убыванию»

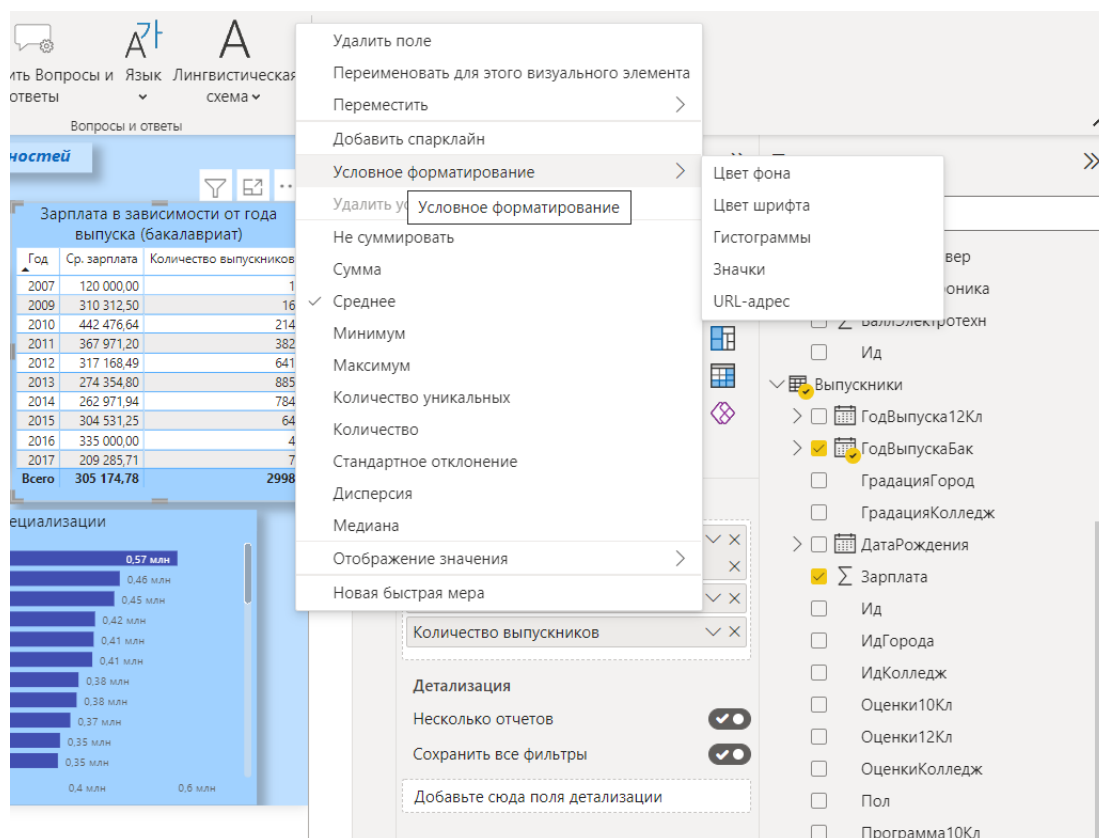


Аналогичную операцию проведем и для гистограммы с группировкой.



Большую среднюю заработную плату имеют выпускники с ученой степенью M.Sc. и получившие диплом по специальности «computer networking» (те специализация в большей мере влияет на зарплату, нежели другие рассмотренные выше показатели).

Чтобы применить условное форматирование, необходимо выбрать элемент визуализации (в нашем случае таблица), затем на панели Визуализации щелкнуть стрелку вниз рядом с полем Ср.Зарплата, которое будет отформатировано и далее выбрать Условное форматирование и необходимый тип.



Отформатируем цвета ячеек количества выпускников в зависимости от правил. Для этого выберем Условное форматирование – Цвет фона и заполним поля в появившемся окне (при выборе варианта Процент границы правила настраиваются как процент от общего диапазона значений от минимума до максимума).

Цвет фона - Количество выпускников

Стиль формата

Правила

Применить

Только значения

Какое поле следует использовать в качестве основы?

Количество Ид

Суммирование

Количество

Правила

↑↓ Обратить порядок цвет...

+ Новое правило

Если значение	>=	0	Процент	и	<=	25	Процент	то	lightblue	↑ ↓ ×
Если значение	>	25	Процент	и	<=	50	Процент	то	lightblue	↑ ↓ ×
Если значение	>	50	Процент	и	<=	75	Процент	то	blue	↑ ↓ ×
Если значение	>	75	Процент	и	<=	100	Процент	то	darkblue	↑ ↓ ×

После применения правила:

Зарплата в зависимости от года выпуска (бакалавриат)		
Год	Ср. зарплата	Количество выпускников
2007	120 000,00	1
2009	310 312,50	16
2010	442 476,64	214
2011	367 971,20	382
2012	317 168,49	641
2013	274 354,80	885
2014	262 971,94	784
2015	304 531,25	64
2016	335 000,00	4
2017	209 285,71	7
Всего	305 174,78	2998

Для форматирования средней зарплаты выберем Условное форматирование – Гистограммы (за 100% здесь также берется наибольшее значение из всех строк).

Гистограммы - Ср. зарплата

Отображение в ячейках горизонтальных столбцов на основе их значений.

☐ Показывать только столбец

Минимум

Самое низкое значение

Введите значение

Максимум

Самое высокое значение

Введите значение

Столбец с положит. зн...

lightblue

Направление столбца

Слева направо

Столбец с отриц. знач.

lightblue

Ось

darkblue

Результат

Зарплата в зависимости от года выпуска (бакалавриат)		
Год	Ср. зарплата	Количество выпускников
2007	20 000,00	1
2009	310 312,50	16
2010	442 476,64	214
2011	367 971,20	382
2012	317 168,49	641
2013	274 354,80	885
2014	262 971,94	784
2015	304 531,25	64
2016	335 000,00	4
2017	209 285,71	7
Всего	305 174,78	2998

По аналогии можно применить форматирование относительно цвета шрифта/добавить значки.

Использование фильтров помогает сократить объем отображаемой информации: так, можно убрать из представления те специализации, количество которых мало по сравнению с остальными (1-5) или же средняя зарплата которых не попадает в определенный диапазон. Для этого нужно воспользоваться разделом Фильтры и применить необходимые критерии вручную, либо используя правила (расширенная фильтрация).

Зарплата в зависимости от года выпуска (бакалавриат)

Год	Ср. зарплата	Количество выпускников
2007	20 000,00	1
2009	310 312,50	16
2010	442 476,64	214
2011	367 971,20	382
2012	317 168,49	641
2013	274 354,80	885
2014	262 971,94	784
2015	304 531,25	64
2016	335 000,00	4
2017	209 285,71	7
Всего	305 174,78	2998

Ср. зарплата по специализации

computer networking	0,57 млн
information & commu...	0,45 млн
chemical engineering	0,42 млн
industrial & producti...	0,41 млн
computer engineering	0,38 млн
electronics & instrum...	0,38 млн
industrial engineering	0,37 млн
telecommunication en...	0,35 млн
civil engineering	0,35 млн
metallurgical engineer...	0,34 млн
mechanical engineering	0,33 млн

Фильтры

Поиск

Фильтры для этого визуального элемента

Специализация
не является biomedical engineering, ceramic e...

Тип фильтра ⓘ
Простая фильтрация

Поиск

- ☐ information science 1
- ☒ information science engineering 18
- ☒ information technology 506
- ☐ instrumentation and control engineering 18
- ☐ instrumentation engineering 3
- ☐ mechanical & production engineering 1
- ☒ mechanical and automation 5

☐ Нужно выбрать один элемент

Ср. зарплата

меньше чем или равно 600000 и больше че...

Показать элементы, когда значение

меньше или равно

600000

☒ И ☐ Или

больше или равно

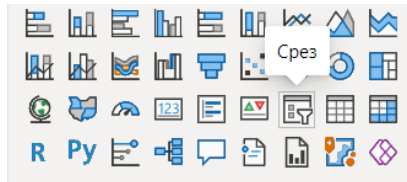
400000

Ср. зарплата по специализации

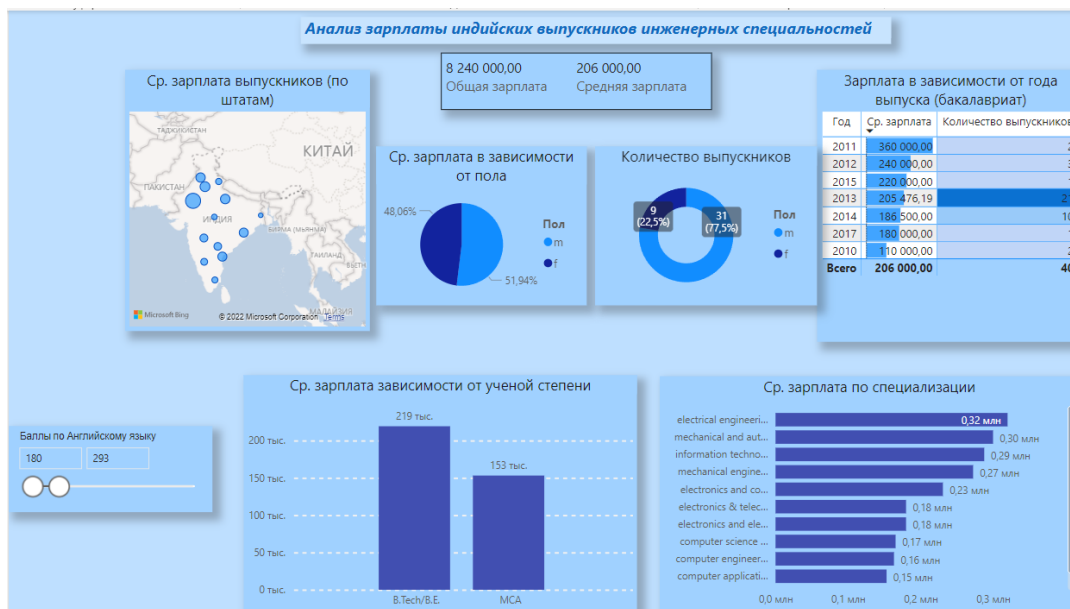
computer networking	0,57 млн
information science	0,40 млн
information & commu...	0,45 млн
chemical engineering	0,42 млн
industrial & producti...	0,41 млн
instrumentation and c...	0,41 млн

В отличие от фильтров, срезы присутствуют в отчете как визуальные элементы и позволяют выбирать значения при анализе отчета. Все остальные визуальные элементы в отчете изменяются в соответствии с диапазоном, который установлен на панели среза.

Для добавления нужно выбрать значок

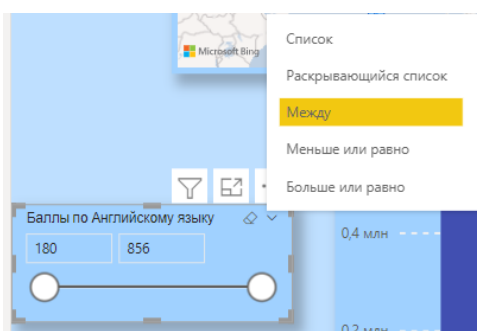


Добавим срез для оценки влияния баллов по английскому языку на среднюю зарплату

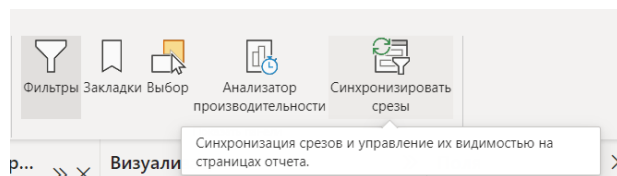


Изменение баллов влияет как на зарплату в целом, так и на структуру распределения специализаций/полученных ученых степеней. Низким баллам соответствуют невысокие средние зарплаты (а ученая степень в основном - бакалавр).

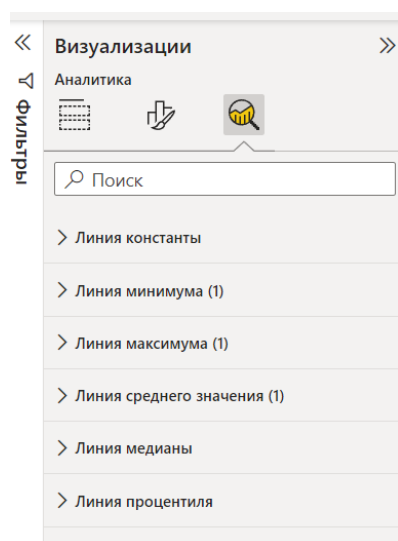
Тип среза можно изменить в раскрывающейся панели



Также доступна опция синхронизации срезов на вкладке Представление (если они располагаются на разных страницах)



Панель Аналитика позволяет добавлять динамические опорные линии для визуальных элементов, доступна из области Визуализации (для элементов карта, круговая диаграмма, кольцевой график, таблица отсутствует). Доступные линии для линейчатой гистограммы:



Добавим для средней зарплаты линии, отмечающие минимум, среднее и максимум.



При изменении баллов по английскому в срезе информация динамически обновляется.



8. Добавить в отчет вычисляемые столбцы, вычисляемые поля (меры).

Создадим вычисляемый столбец, отражающий информацию о возрасте выпускников

Поля

Поиск

Балл_AMCAT

Создать меру

Создать столбец

Новая быстрая мера

Обновить данные

Изменить запрос

Управление связями

Дополнительное обновление

Управление агрегированием

Переименовать

Удалить из модели

Скрыть

Отметить как таблицу дат

Показать скрытые

Отменить скрытие всех

Свернуть все

Развернуть все

ГодВыпуска12Кл

ГодВыпускаБак

ГрадацияГород

ГрадацияКолле...

ДатаРождения

Зарплата

Ид

ИдГорода

ИдКолледж

Оценки10Кл

Оценки12Кл

ОценкиКолледж

Пол

Программа10Кл

Программа12Кл

Специализация

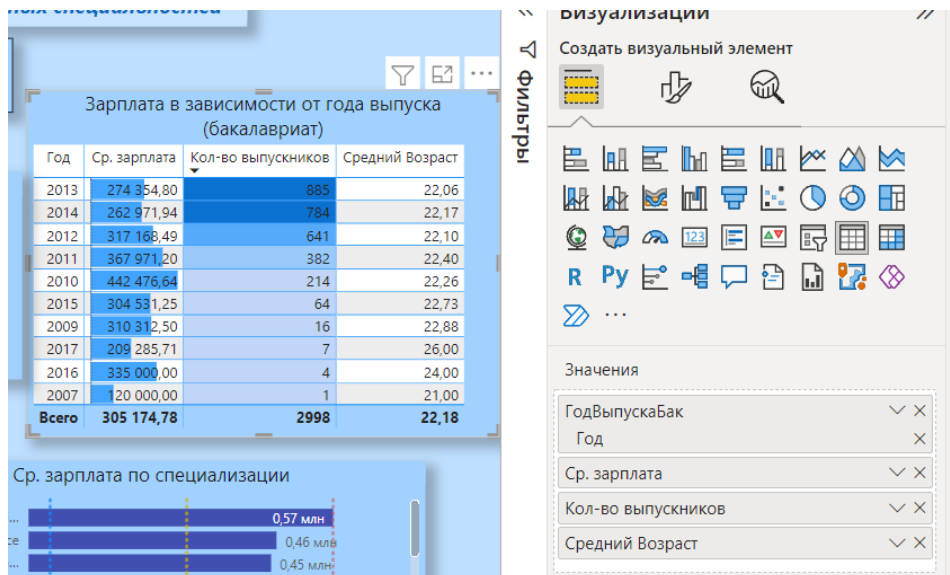
Структура

Форматирование

Свойства

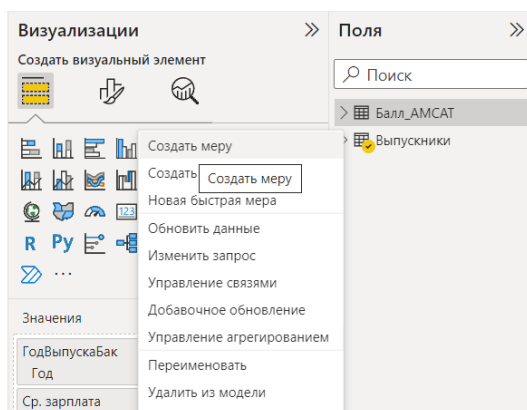
1 Возраст = YEAR('Выпускники'[ГодВыпускаБак]) - YEAR('Выпускники'[ДатаРождения])

Отобразим эту информацию в таблице



Вычисляемые столбцы используются для формирования значений в контексте каждой отдельной строки, а меры агрегируют эти значения для всей таблицы.

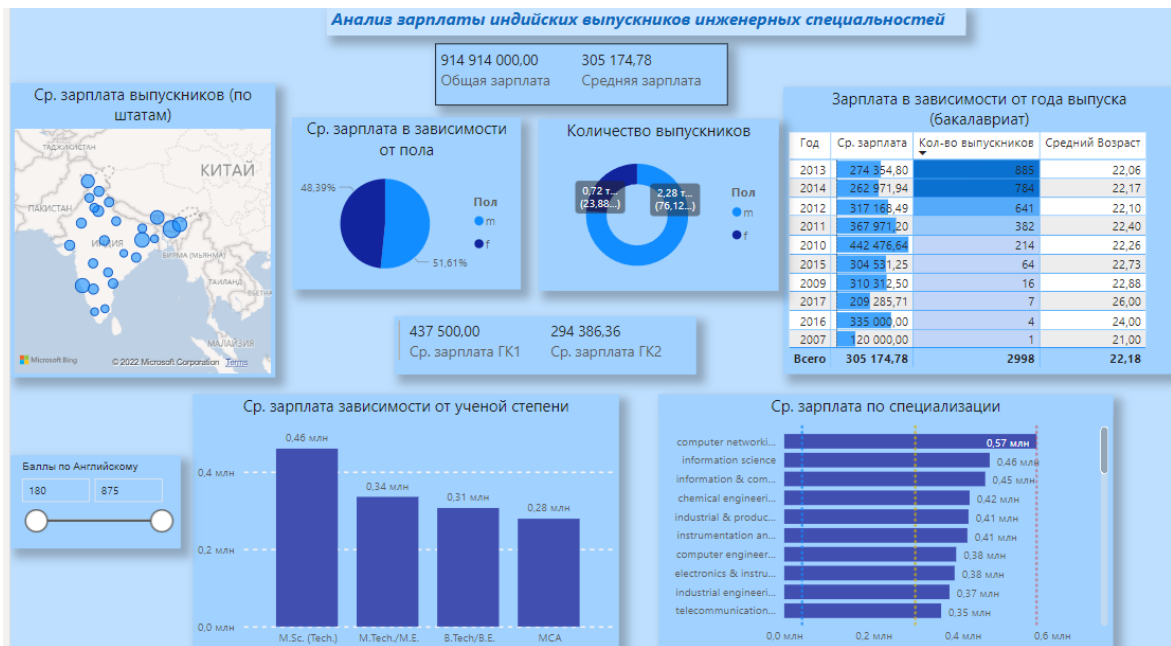
Создадим 2 меры, отображающие средние зарплаты выпускников в зависимости от градации колледжа (1 считается лучше).



```
1 Ср. зарплата ГК1 = CALCULATE (
2 AVERAGE('Выпускники'[Зарплата]),
3 'Выпускники'[ГрадацияКолледж] = "1"
4 )
```

```
1 Ср. зарплата ГК2 = CALCULATE (
2 AVERAGE('Выпускники'[Зарплата]),
3 'Выпускники'[ГрадацияКолледж] = "2"
4 )
```

Отообразим информацию в карточке



Видно, что средняя зарплата у выпускников колледжа 1 категории выше, скорее всего это связано с более качественным преподаванием.

Также можно добавить меру с общим расчетом количества выпускников той или иной категории.

Выпускники ГК1 = COUNTROWS (FILTER ('Выпускники', 'Выпускники'[ГрадацияКолледж] = "1"))

Выпускники ГК2 = COUNTROWS (FILTER ('Выпускники', 'Выпускники'[ГрадацияКолледж] = "2"))



9. Продемонстрировать использование параметра (What if) для варьирования и анализа изменения результатов.

Для демонстрации использования параметра What if (вкладка Моделирование – создание параметра) оценим заработную плату с учетом некоторого процента налога.

Параметр What-if

Имя
Параметр

Тип данных
Десятичное число

Минимум
0

Максимум
0,30

Увеличение
0,01

По умолчанию

☒ Добавить срез на страницу

OK Отмена

После этого на странице добавится срез. Также необходимо создать новую меру для отображения результатов, которую потом можно также отобразить в таблице.

Ср ЗП с налогом = $AVERAGE('Выпускники'[Зарплата]) - AVERAGE('Выпускники'[Зарплата]) * \text{Налог, \%}$ [Значение Налог, %]

Изменения при увеличении налога



10. Осуществить экспорт отчетов в PDF-файл.

Экспорт в PDF-файл осуществляется при последовательном переходе «Файл»-«Экспортировать»-«Экспортировать в PDF»

←

Создать

Открыть отчет

Сохранить

Сохранить как

Получить данные

Импортировать

Экспортировать

Экспортировать

Шаблон Power BI

Экспортировать в PDF

Экспортировать текущий отчет в PDF

Итог

Анализ зарплаты индийских выпускников инженерных специальностей

914 914 000,00

305 174,78

Общая зарплата

Средняя зарплата

Ср. зарплата выпускников (по штатам)

Ср. зарплата в зависимости от пола

48,39%

51,61%

Пол

m

f

Количество выпускников

0,72 тыс. (23,88%)

2,28 тыс. (76,12%)

Пол

m

f

Зарплата в зависимости от года выпуска (бакалавриат)

Год	Ср. зарплата	Выпускников	Ср. Возраст	Ср ЗП с налогом
2013	274 354,80	885	22,06	274 354,80
2014	262 971,94	784	22,17	262 971,94
2012	317 168,49	641	22,10	317 168,49
2011	367 971,20	382	22,40	367 971,20
2010	442 476,64	214	22,26	442 476,64
2015	304 531,25	64	22,73	304 531,25
2009	310 312,50	16	22,88	310 312,50
2017	209 285,71	7	26,00	209 285,71
2016	335 000,00	4	24,00	335 000,00
2007	120 000,00	1	21,00	120 000,00
Всего	305 174,78	2998	22,18	305 174,78

Ср. зарплата ГК1

294 386,36

Ср. зарплата ГК2

226

Выпускников ГК1

2772

Выпускников ГК2

Баллы по Английскому

180

875

Набор, %

0,00

Ср. зарплата зависимости от ученой степени

Ученая степень	Ср. зарплата
M.Sc. (Tech.)	0,46 млн
M.Tech./M.E.	0,34 млн
B.Tech/B.E.	0,31 млн
MCA	0,28 млн

Ср. зарплата по специализации

Специализация	Ср. зарплата
computer network...	0,57 млн
information science	0,46 млн
information & com...	0,45 млн
chemical engineeri...	0,42 млн
industrial & produc...	0,41 млн
instrumentation an...	0,41 млн
computer engineer...	0,38 млн
electronics & instru...	0,38 млн
industrial engineeri...	0,37 млн