

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российский экономический университет имени Г.В. Плеханова»

Высшая школа кибертехнологий, математики и статистики

Направление Бизнес-информатика

Кафедра Информатики

ОТЧЕТ

по дисциплине: «Хранилища данных и системы Business Intelligence»
на тему: «Построение витрины данных и системы аналитической отчетности
для бизнес-анализа продаж категорий велосипедов»

Выполнила
обучающиеся группы
15.11Д-БИЦТ09/216
очной формы обучения
Высшей школы кибертехнологий,
математики и статистики
Крабу Кира Сергеевна

Научный руководитель:
Дьяконова Л.П., к.н.,
доцент кафедры информатики

Москва 2024

Оглавление

Краткая постановка задачи	3
Список аналитических отчетов	3
Логическая схема Витрины данных	5
Подготовка данных к загрузке в ВІ-платформу	5
Построение схемы данных в ВІ-платформе.....	10
Создание системы аналитической отчетности средствами ВІ-платформы	12
Краткие выводы.....	26

Краткая постановка задачи

Данный отчет выполнен на тему «Построение витрины данных и системы аналитической отчетности для бизнес-анализа продаж категорий товаров» по данным файла «SAP Велосипеды».

Файл Excel с данными о велосипедах содержит в себе такие данные о продажах: Order ID, Date, Sales Agent Last Name, Sales Agent First Name, Customer, Customer Segment, Country, Latitude, Longitude, Customer Status, Product, Product Type, No Customer Meetings, Units Sold, Order Value.

Будущий дашборд и система аналитической отчетности могут быть полезны для руководителей компании, менеджеров по продажам, маркетологов и аналитиков. Он позволяет отслеживать динамику продаж товаров различных категорий, выявлять тенденции и паттерны, выявлять наиболее популярные товары и категории, а также оценивать эффективность маркетинговых кампаний и принимать обоснованные решения для оптимизации продаж.

Список аналитических отчетов

Таблица 1. Аналитические отчеты

№	Название	Список полей	Уровень агрегации	Функции агрегации	Период отчёта	Вид отчёта
1	Продажи по категориям товаров	Product type Product Units Sold Order Value	Product type Product	Сумма	2 года	Гистограмма
2	Сегментация рынка	Product type Product Customer Segment Units Sold Order Value	Customer Segment Product type Product	Сумма	2 года	Гистограмма

3	Географическое распределение продаж	Product type Product Country Latitude Longitude Units Sold Order Value	Country Product type Product	Сумма	2 года	Тепловая карта
4	Статус покупателя	Product type Product Customer Status Units Sold Order Value	Customer Status Product type Product	Сумма	2 года	Столбчатая диаграмма
5	Динамика продаж категорий товаров	Product type Product Date Units Sold Order Value	Date Product type Product	Сумма	2 года	Линейный график
6	Информационная панель					

Логическая схема Витрины данных

Логическая схема нашей базы данных – «звездочка».

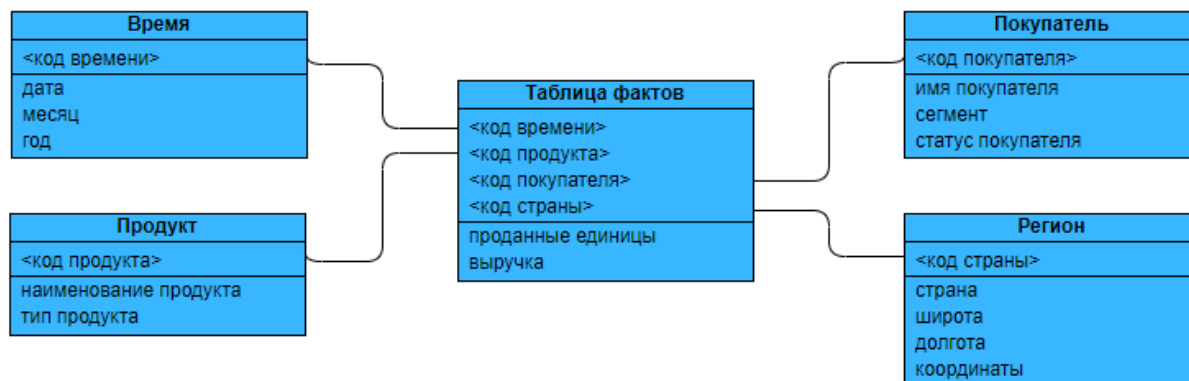


Рисунок 1. Логическая схема БД

Подготовка данных к загрузке в BI-платформу

Для построения дашборда мы будем загружать данные в Yandex DataLens – облачную BI-платформу. Для этого данные сначала нужно подготовить.

Во-первых, изначальный набор данных нужно разделить на пять таблиц в соответствии со схемой выше.

Во-вторых, нужно поработать с полями и «удалить» те, которые для нас не несут никакой информации. В нашем случае это поля: Order, Sales Agent Last Name, Sales Agent First Name, No Customer Meetings. Они не будут использованы ни в какой таблице.

В-третьих, нужно будет добавить необходимые столбцы, например, столбцы с кодами данных. Все ID делались по одному и тому же принципу (кроме ID времени).

Для таблицы «Время» DateID было создано с помощью функции СЦЕПИТЬ месяц и год, которые предварительно были «вытащены» с основной даты функциями ГОД и МЕСЯЦ.

DateID	Month	Year	Date
52019	5	2019	29.05.2019
52019	5	2019	29.05.2019
52019	5	2019	29.05.2019
52019	5	2019	29.05.2019
52019	5	2019	29.05.2019
52019	5	2019	29.05.2019
52019	5	2019	29.05.2019
52019	5	2019	29.05.2019
42019	4	2019	29.04.2019
42019	4	2019	29.04.2019
42019	4	2019	29.04.2019
42019	4	2019	29.04.2019
42019	4	2019	29.04.2019
42019	4	2019	29.04.2019
42019	4	2019	29.04.2019
42019	4	2019	29.04.2019
42019	4	2019	29.04.2019
32019	3	2019	29.03.2019
32019	3	2019	29.03.2019

Рисунок 2. Часть таблицы «Время»

Остальные ID было создано с помощью сводных таблиц и функции ВПР. Построив сводную таблицу для нужных данных, ID были распределены по уникальным экземплярам, и далее присвоены данным на основном листе. Для таблиц «Продукт» и «Покупатель» никаких дополнительных столбцов (кроме ID) создано не было.

Названия строк		Product	id
C900 BIKE	1	C900 BIKE	1
C950 BIKE	2	C950 BIKE	2
C990 Bike	3	C990 Bike	3
eBike E148	4	eBike E148	4
M525 BIKE	5	M525 BIKE	5
M550 BIKE	6	M550 BIKE	6
R100 BIKE	7	R100 BIKE	7
R200 BIKe 4s	8	R200 BIKe	8
R300 Bike	9	R300 Bike	9

Рисунок 3. Сводная таблица для «Продукта»

Далее для этой же таблицы и создании столбца координат нужно было сцепить широту и долготу (рисунок 5).

В итоге получаются такие таблицы (вернее сказать листы, данные с которых и будут загружены в BI-платформу.

	A	B	C	D
1	ProductID	Product	Product Type	
2	B1	C900 BIKE	Mountain	
3	B2	C950 BIKE	Mountain	
4	B3	C990 Bike	Racing	
5	B4	eBike E148	Mountain	
6	B5	M525 BIKE	Mountain	
7	B6	M550 BIKE	Mountain	
8	B7	R100 BIKE	Racing	
9	B8	R200 BIKe 4s	Racing	
10	B9	R300 Bike	Racing	
11				

Рисунок 7. Таблица «Продукт»

	A	B	C	D	E
1	Customer	Customer	Customer	Customer Status	
2	C1	100 Rides	Wholesale	Prior Customer	
3	C10	4mation	Direct	Prior Customer	
4	C100	APLus Wh	Retail	Prospect	
5	C1000	Hopper Bi	Direct	Current Customer	
6	C1001	Horizon C	Direct	Current Customer	
7	C1002	Hornet Bil	Retail	Prospect	
8	C1003	Host Spee	Retail	Prospect	
9	C1004	Hot Whee	Wholesale	Current Customer	
10	C1005	Hubbub	Retail	Prospect	
11	C1006	Hubert's	Retail	Prospect	
12	C1007	Human Co	Direct	Prior Customer	
13	C1008	Humming	Retail	Prior Customer	
14	C1009	Hunter Bic	Wholesale	Current Customer	
15	C101	Apogee W	Direct	Prior Customer	
16	C1010	Hwel	Direct	Prior Customer	
17	C1010	Hwel	Wholesale	Prospect	
18	C1011	Iconic Spe	Wholesale	Prior Customer	
19	C1012	Ignition S	Retail	Prior Customer	
20	C1013	Illuminate	Wholesale	Current Customer	
21	C1014	Immediat	Direct	Prior Customer	
22	C1015	Immortal	Wholesale	Current Customer	
23	C1016	Impact	Direct	Prospect	

Рисунок 8. Таблица «Покупатель»

	A	B	C	D	E	F
1	CountryID	Country	Latitude	Longitude	Coordinates	
2	CO1	Afghanistan	33.93911	67.709953	[33.93911,67.709953]	
3	CO10	Argentina	-38.416097	-63.61667	[-38.416097,-63.61667]	
4	CO100	Israel	31.046051	34.851612	[31.046051,34.851612]	
5	CO101	Italy	41.87194	12.56738	[41.87194,12.56738]	
6	CO102	Jamaica	18.109581	-77.297508	[18.109581,-77.297508]	
7	CO103	Japan	36.204824	138.25292	[36.204824,138.25292]	
8	CO104	Jordan	30.585164	36.238414	[30.585164,36.238414]	
9	CO105	Kazakhstan	48.019573	66.923684	[48.019573,66.923684]	
10	CO106	Kenya	-0.023559	37.906193	[-0.023559,37.906193]	
11	CO107	Kiribati	-3.370417	-168.73403	[-3.370417,-168.73403]	
12	CO108	Kuwait	29.31166	47.481766	[29.31166,47.481766]	
13	CO109	Kyrgyzstan	41.20438	74.766098	[41.20438,74.766098]	
14	CO11	Armenia	40.069099	45.038189	[40.069099,45.038189]	
15	CO110	Latvia	56.879635	24.603189	[56.879635,24.603189]	
16	CO111	Lebanon	33.854734	35.862285	[33.854734,35.862285]	

Рисунок 9. Таблица «Регион»

	A	B	C	D
1	DateID	Date	Month	Year
2	102017	27.10.2017	10	2017
3	102018	27.10.2018	10	2018
4	112017	27.11.2017	11	2017
5	112018	27.11.2018	11	2018
6	12018	27.01.2018	1	2018
7	12019	27.01.2019	1	2019
8	122017	27.12.2017	12	2017
9	122018	27.12.2018	12	2018
10	22018	26.02.2018	2	2018
11	22019	26.02.2019	2	2019
12	32018	29.03.2018	3	2018
13	32019	29.03.2019	3	2019
14	42018	29.04.2018	4	2018
15	42019	29.04.2019	4	2019
16	52018	29.05.2018	5	2018
17	52019	29.05.2019	5	2019
18	62017	29.06.2017	6	2017
19	62018	29.06.2018	6	2018
20	72017	29.07.2017	7	2017
21	72018	29.07.2018	7	2018
22	82017	29.08.2017	8	2017
23	82018	29.08.2018	8	2018

Рисунок 10. Таблица «Время»

	A	B	C	D	E	F	
1	DateID	CustomerID	CountryID	ProductID	Units Sold	Order Value	
2	102017	C1036	CO121	B4	201	6132	
3	102017	C104	CO167	B2	113	3099	
4	102017	C1052	CO48	B9	575	13613	
5	102017	C1062	CO66	B1	132	2523	
6	102017	C1145	CO57	B4	75	2643	
7	102017	C1147	CO147	B7	267	4455	
8	102017	C1158	CO44	B8	368	12311	
9	102017	C1189	CO214	B2	20	1511	
10	102017	C1205	CO14	B4	77	4221	
11	102017	C1213	CO159	B5	419	13862	
12	102017	C1216	CO65	B1	315	11146	
13	102017	C1240	CO19	B7	189	5734	
14	102017	C1275	CO185	B7	81	2402	
15	102017	C131	CO199	B8	147	3208	
16	102017	C1313	CO5	B5	112	4923	
17	102017	C1329	CO10	B2	15	4641	
18	102017	C1397	CO222	B6	90	1740	
19	102017	C1447	CO155	B9	84	3407	
20	102017	C1451	CO180	B7	247	7013	
21	102017	C1456	CO184	B6	429	7870	
22	102017	C1459	CO66	B7	368	10770	

Рисунок 11. Таблица фактов

Таблица фактов (рисунок 11) также составлена с помощью сводной таблицы.

Построение схемы данных в BI-платформе

Теперь данные готовы к загрузке в Yandex DataLens. Для начала создаем подключение и загружаем туда все полученные таблицы. Само подключение называется «Велосипеды от SAP», а таблицы загружены в формате «.xlsx».

Файлы					Поиск по столбцу	
+ Загрузить файлы						
					DateID	Date
					#	Month
					#	Year
Время.xlsx - Date	...				102017	2017-10-27T00:00:00
					102018	2018-10-27T00:00:00
Попкупатель.xlsx - Customer	...				112017	2017-11-27T00:00:00
					112018	2018-11-27T00:00:00
Продукт.xlsx - Product	...				12018	2018-01-27T00:00:00
					12019	2019-01-27T00:00:00
Регион.xlsx - Country	...				122017	2017-12-27T00:00:00
					122018	2018-12-27T00:00:00
Таблица фактов.xlsx - Табл...	...				22018	2018-02-26T00:00:00
					22019	2019-02-26T00:00:00
					32018	2018-03-29T00:00:00
					32019	2019-03-29T00:00:00
					42018	2018-04-29T00:00:00
					42019	2019-04-29T00:00:00

Рисунок 12. Загруженные в подключение таблицы

Схема данных такова:

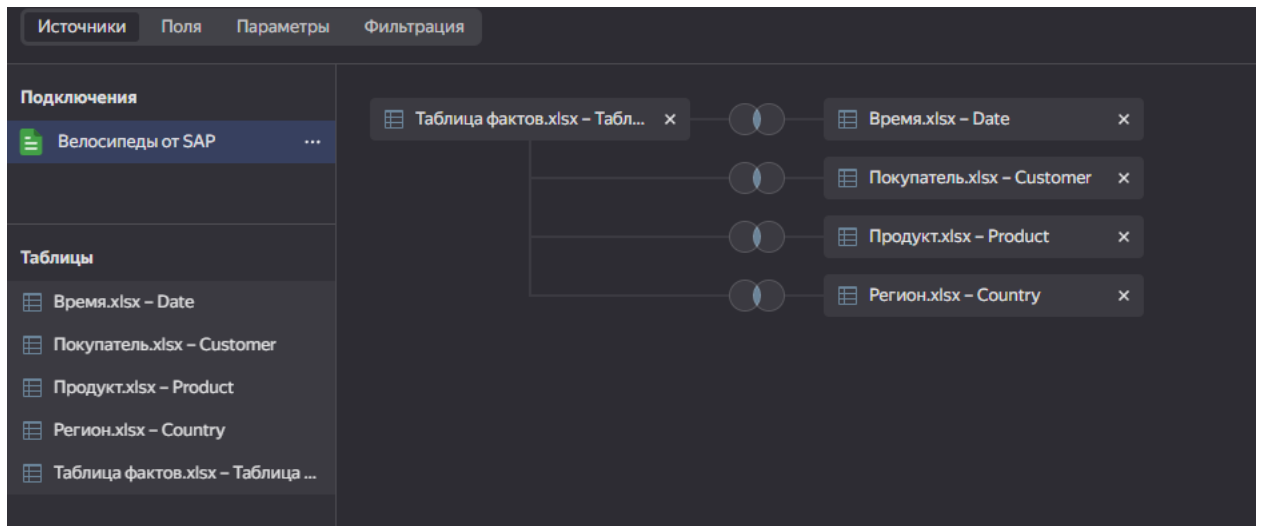


Рисунок 13. Схема данных в Yandex DataLens

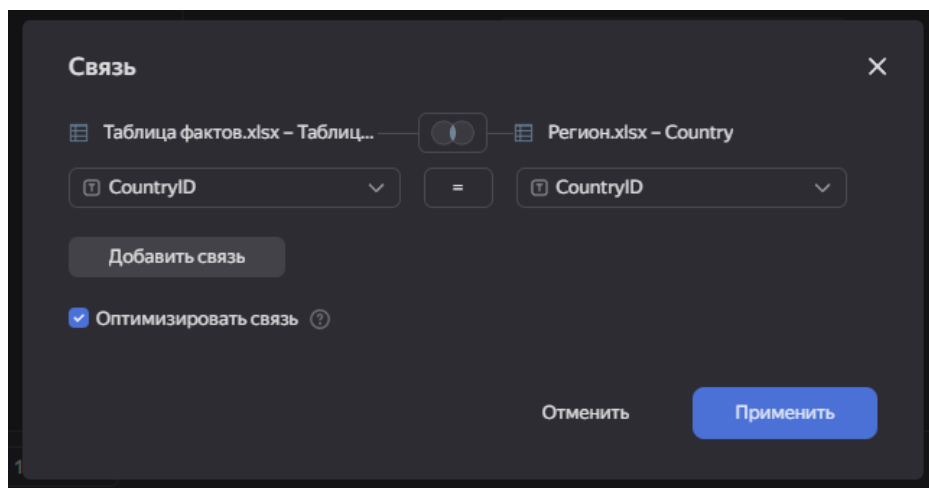


Рисунок 14. Связь между таблицей фактов и «Регионами»

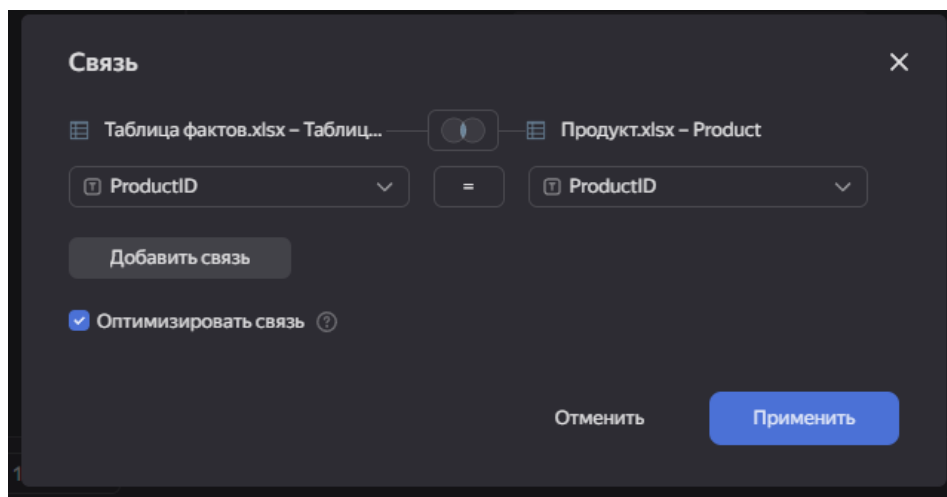


Рисунок 15. Связь между таблицей фактов и «Продуктами»

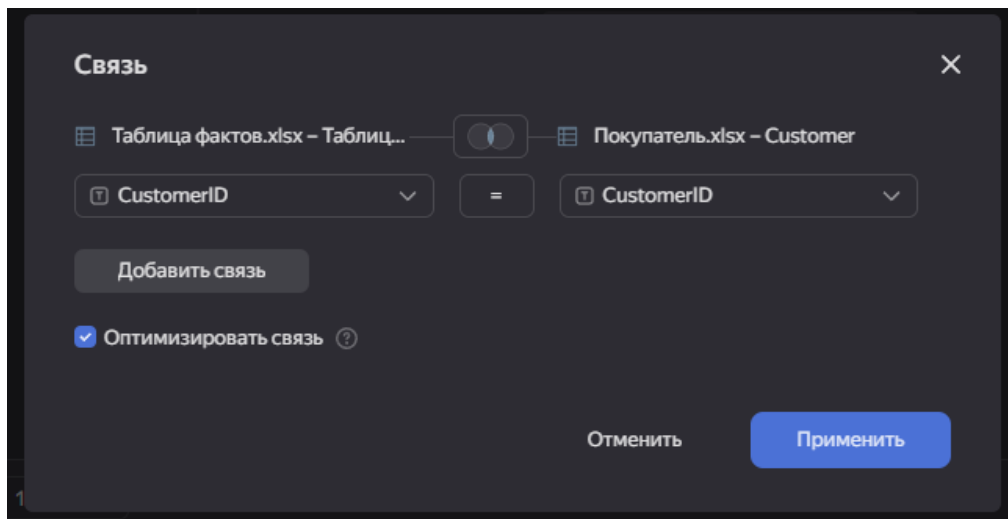


Рисунок 16. Связь между таблицей фактов и «Покупателями»

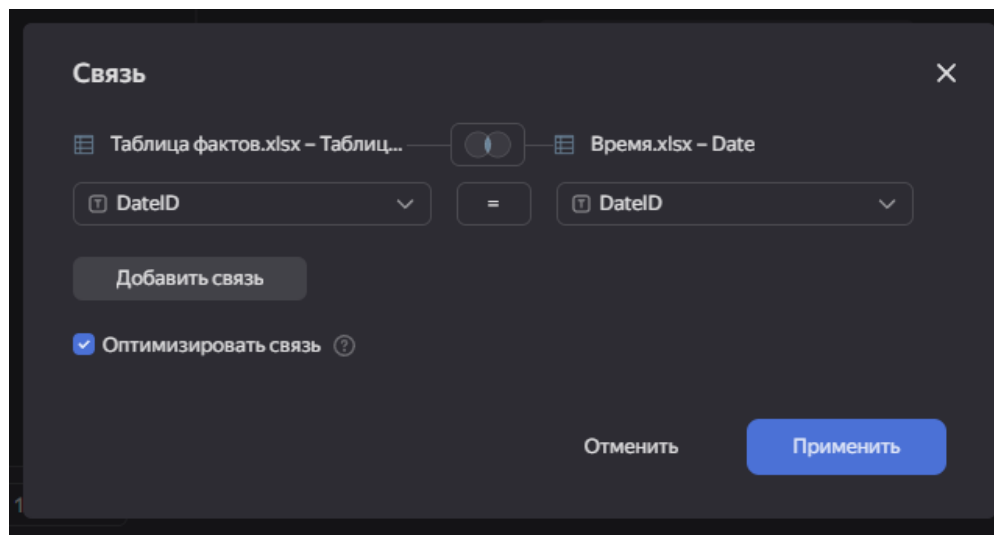


Рисунок 17. Связь между таблицей фактов и «Временем»

Создание системы аналитической отчетности средствами BI-платформы

После того, как подключение было создано, была сформирована схема данных, можно приступить к работе с полями и непосредственно созданию датасета. Изначально удалим создавшиеся дубликаты ID.

19	CountryID (1)	xlsx - Регион.countryid	☐ Строка	Нет
20	ProductID (1)	xlsx - Продукт.productid	# Целое число	Нет
21	CustomerID (1)	xlsx - Покупатель.customerid	# Целое число	Нет
22	DateID (1)	xlsx - Время.dateid	☐ Строка	Нет

Рисунок 18. Дубликаты полей ID

Далее удалим поля «Latitude» и «Longitude», оставив поле «Coordinates» и присвоив ему тип «Геоточка» (по отдельности «Latitude» и «Longitude» не несут для платформы никакой информации). Также поступим с полями «Month» и «Year», оставив «Date» с типом данных «Дата и время».

8	Latitude	xlsx – Регион.latitude	<input checked="" type="checkbox"/> Геоточка	Нет
9	Longitude	xlsx – Регион.longitude	<input checked="" type="checkbox"/> Геоточка	Нет
10	Coordinates	xlsx – Регион.coordinates	<input checked="" type="checkbox"/> Геоточка	Нет

Рисунок 19. Поля «Latitude», «Longitude», «Coordinates»

14	Month	xlsx – Время.month	# Целое число	Нет
15	Year	xlsx – Время.year	# Целое число	Нет
16	Date	xlsx – Время.date	<input checked="" type="checkbox"/> Дата и время	Нет

Рисунок 20. Поля «Month», «Year», «Date»

По итогу получается 14 строк, в которых тип агрегации «Сумма» указан у показателей Units Sold, Order Value.

#	Имя ↓	Источник поля ↓	Тип ↓	Агрегация ↓
1	DateID	xlsx – Таблица фактов.dateid	Строка	Нет
2	CustomerID	xlsx – Таблица фактов.customerid	Строка	Нет
3	CountryID	xlsx – Таблица фактов.countryid	Строка	Нет
4	ProductID	xlsx – Таблица фактов.productid	Строка	Нет
5	Units Sold	xlsx – Таблица фактов.units_sold	# Целое число	Сумма
6	Order Value	xlsx – Таблица фактов.order_value	# Целое число	Сумма
7	Date	xlsx – Date.date	Дата и время	Нет
8	Customer	xlsx – Customer.customer	Строка	Нет
9	Customer Segment	xlsx – Customer.customer_segment	Строка	Нет
10	Customer Status	xlsx – Customer.customer_status	Строка	Нет
11	Product	xlsx – Product.product	Строка	Нет
12	Product Type	xlsx – Product.product_type	Строка	Нет
13	Country	xlsx – Country.country	Строка	Нет
14	Coordinates	xlsx – Country.coordinates	Геоточка	Нет

Рисунок 21. Поля датасета

Далее были созданы несколько чартов в соответствии со списком отчетов, представленным выше.

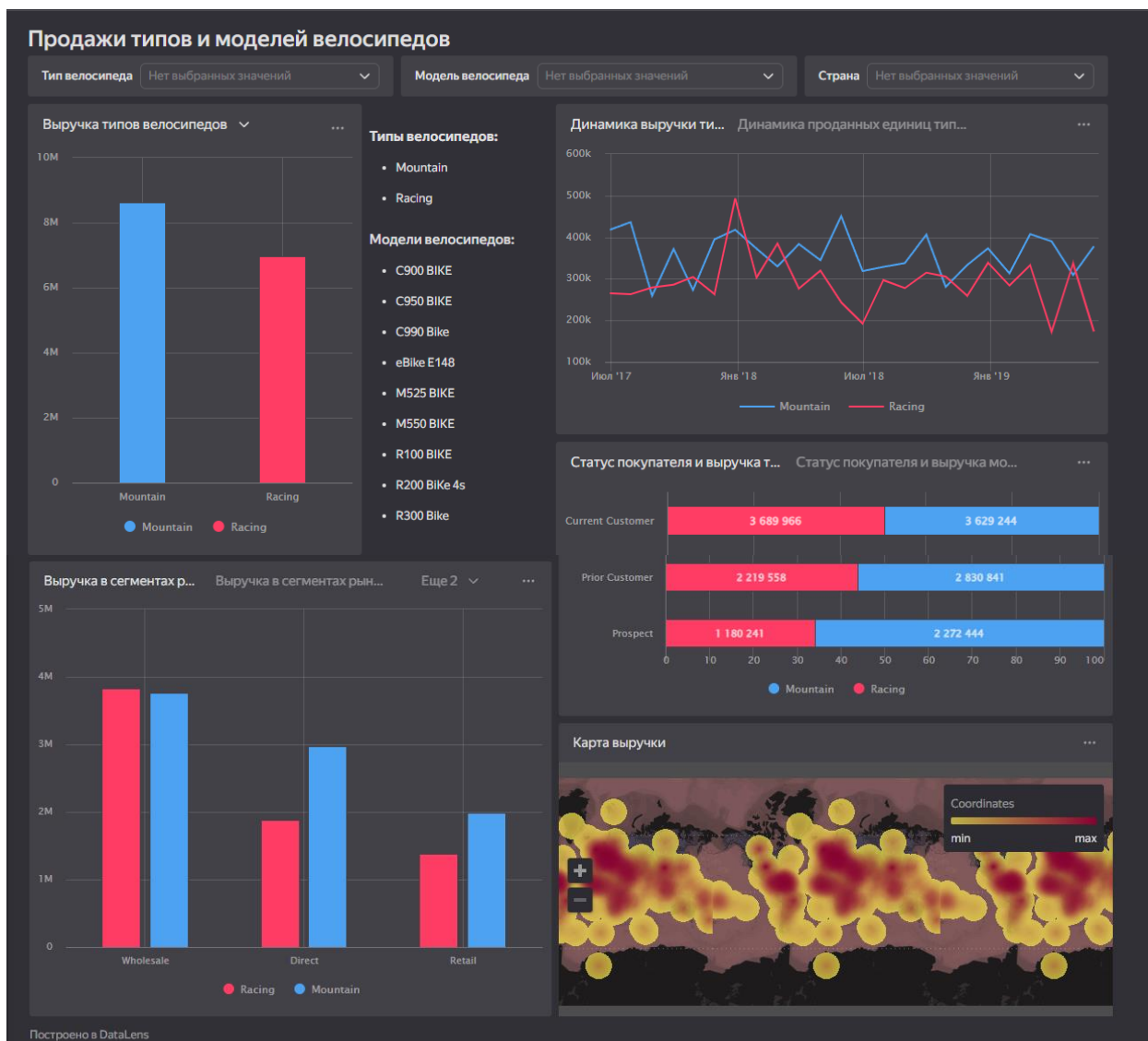


Рисунок 22. Информационная панель

Начнем описывать дашборд с того, что на нем представлено 5 чартов (4 из них могут изменяться в соответствии с целью), 3 селектора: по типу велосипедов, по из моделям и странам. Для большего понимания на дашборде представлены какие есть типы и какие есть модели. Селекторы по типу и по моделям влияют на друг друга (выбирая тип Mountain, в селекторе по моделям для выбора будут представлены только модели для гор), а селектор по странам не «подвержен их влиянию». Также селекторы по типам влияют на данные, связанные с типами, селектор по моделям — на связанные с моделями.

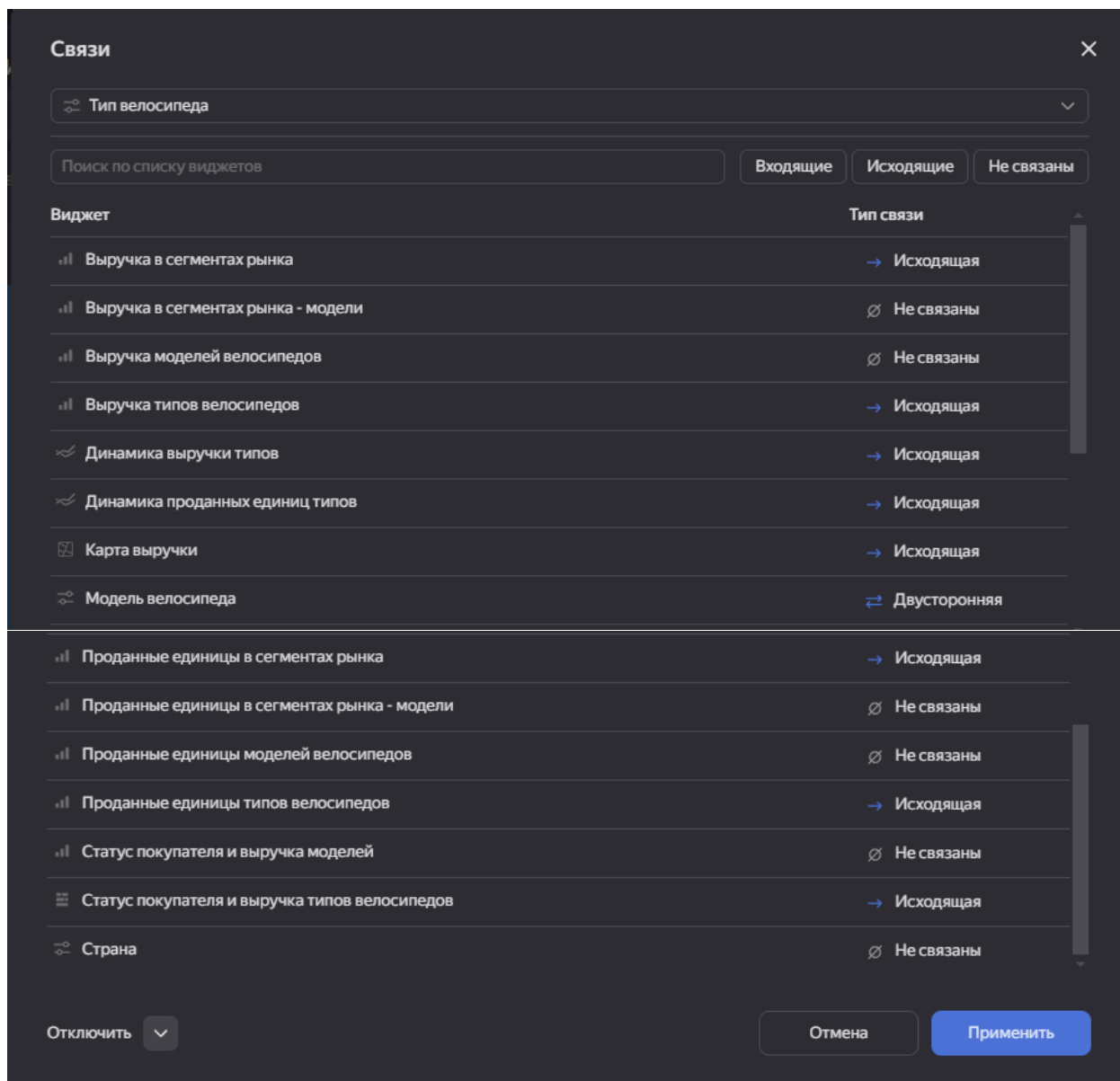


Рисунок 23. Настройки селектора по типу велосипеда

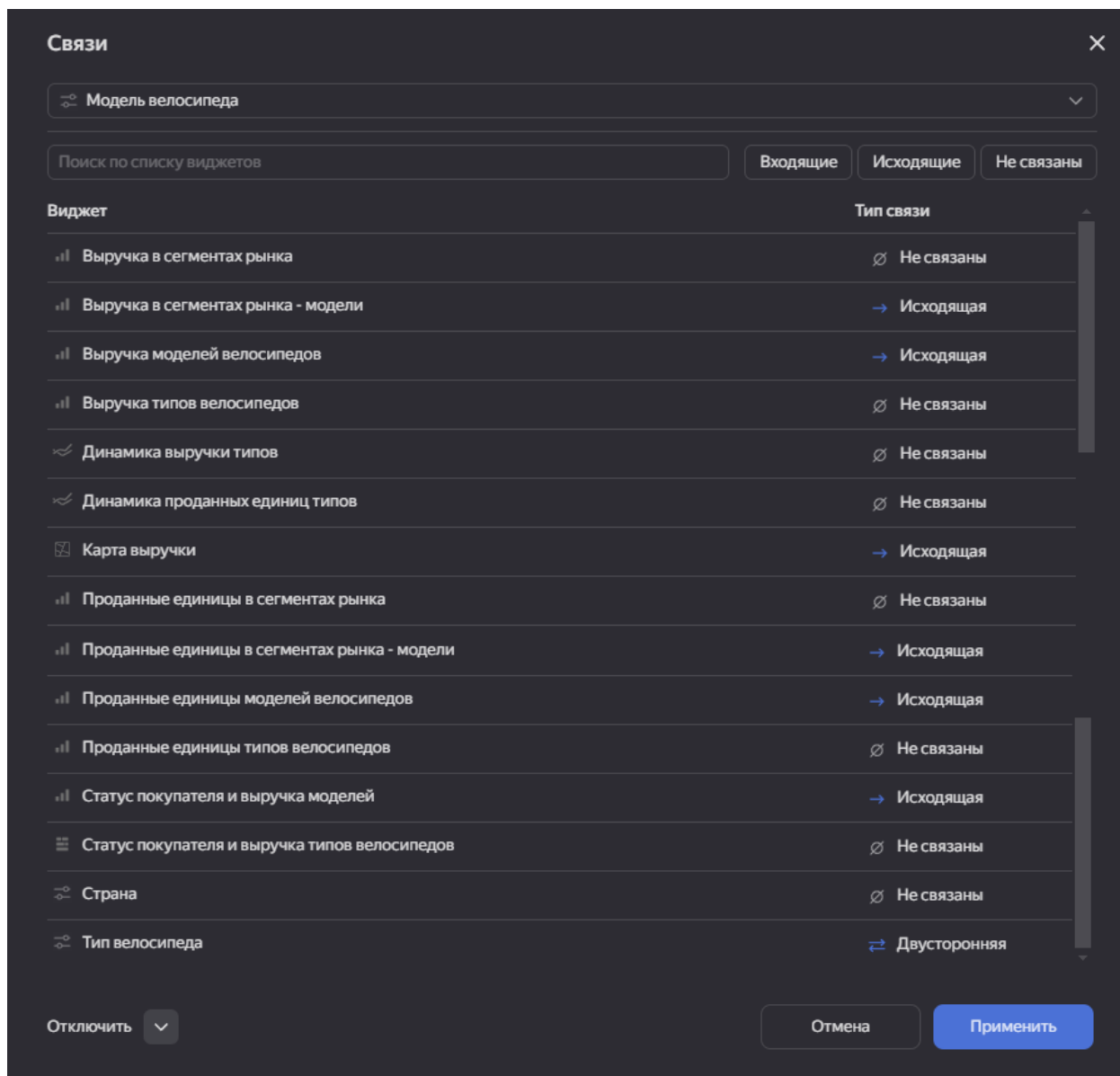


Рисунок 24. Настройки селектора по моделям велосипедов

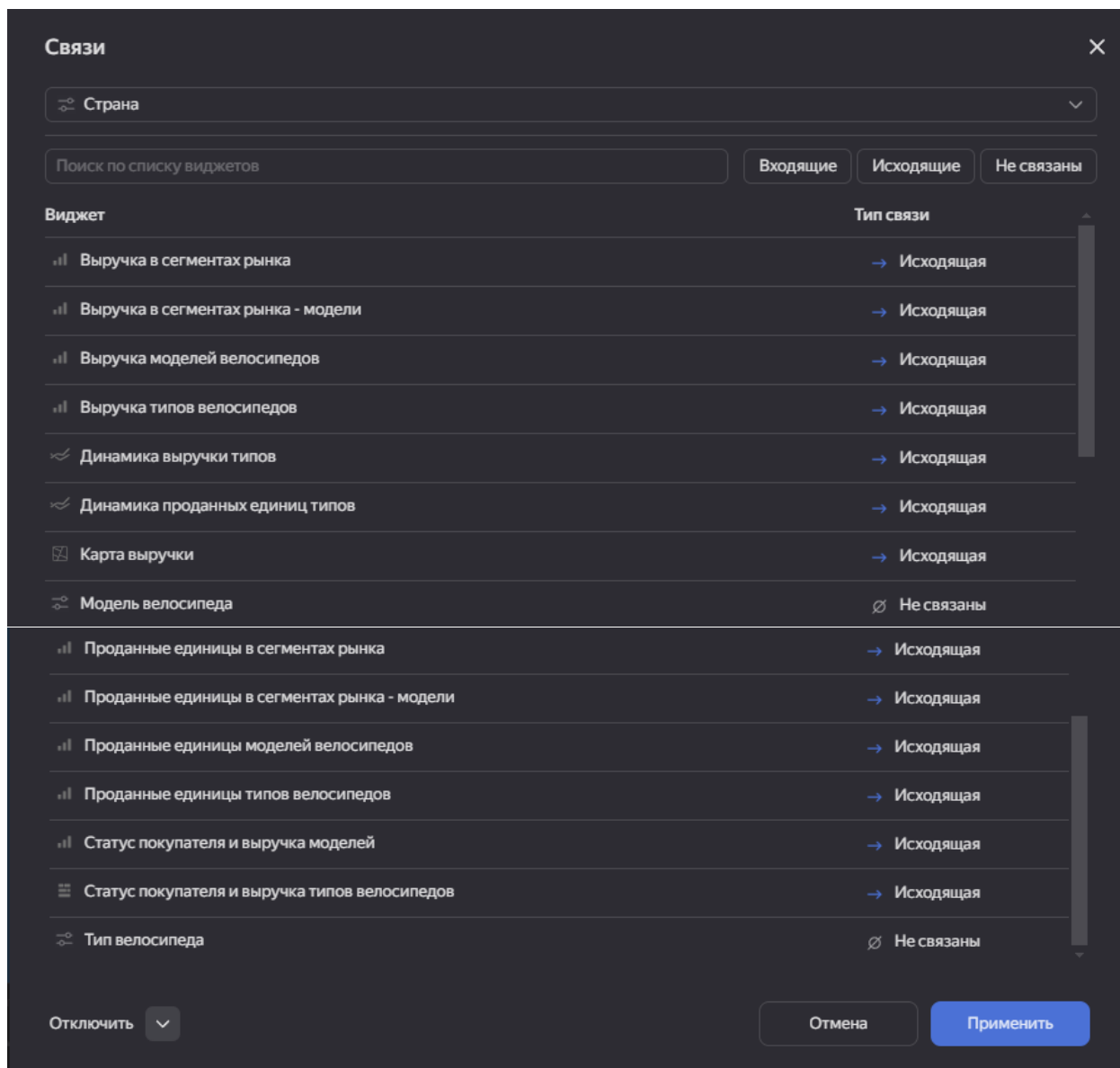


Рисунок 25. Настройки селектора по странам

Рассмотрим поближе отчеты и чарты. Первыми будут чарты, которые несут в себе информацию о продажах по категориям товаров.

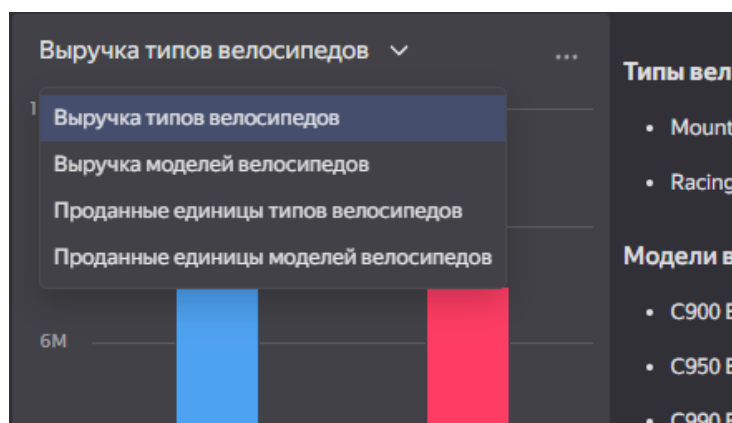


Рисунок 26. Чарты на выбор по информации о продажах

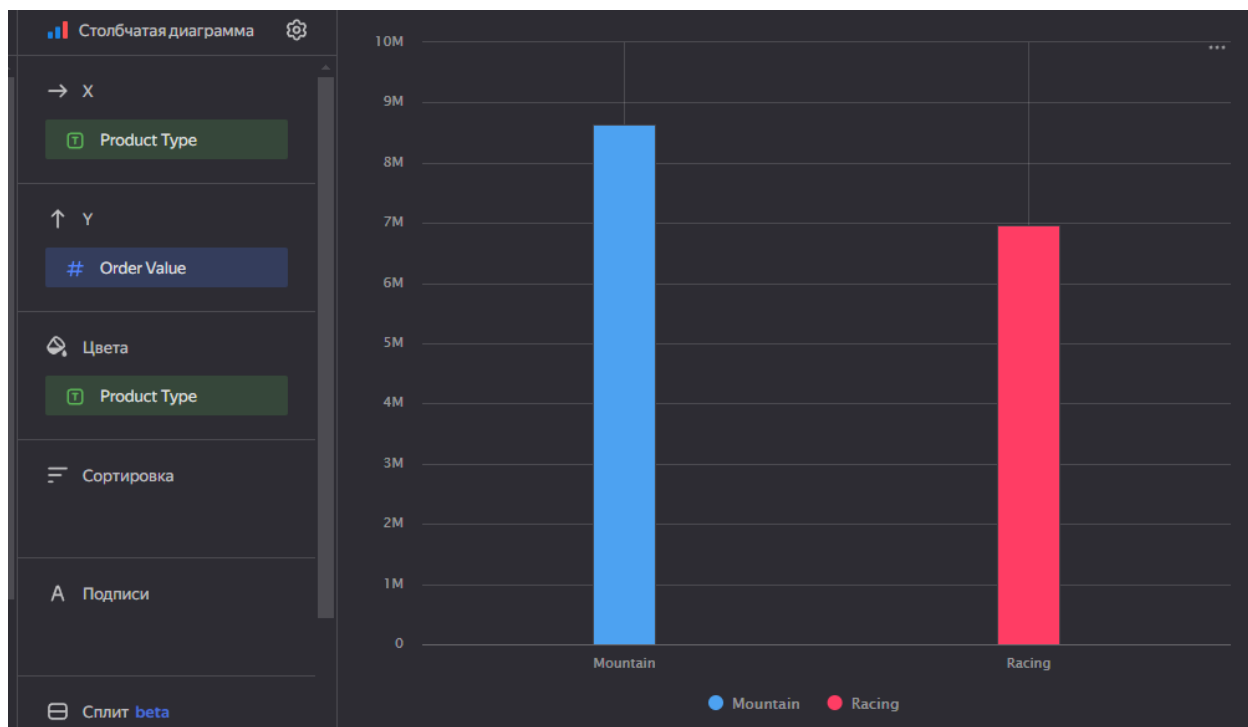


Рисунок 27. Выручка типов велосипедов

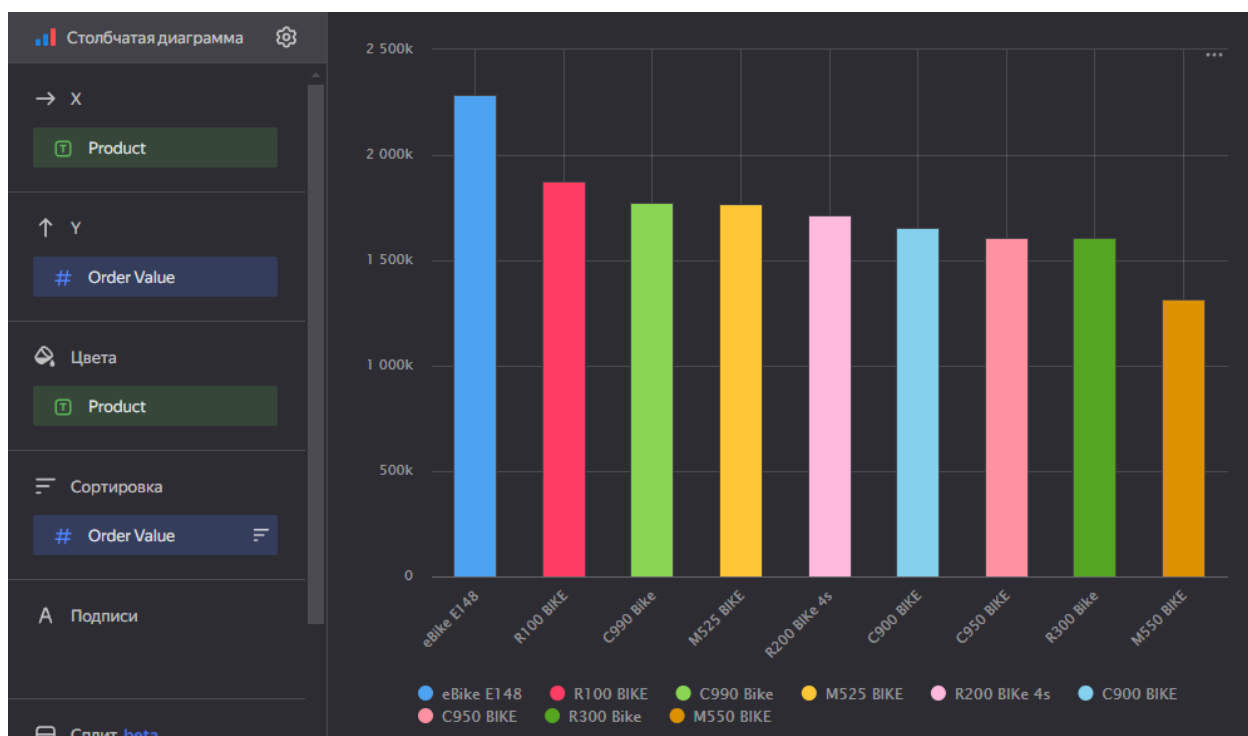


Рисунок 28. Выручка моделей велосипедов



Рисунок 29. Проданные единицы типов велосипедов



Рисунок 30. Проданные единицы моделей велосипедов

Перейдем ко второй теме: сегментация рынка.

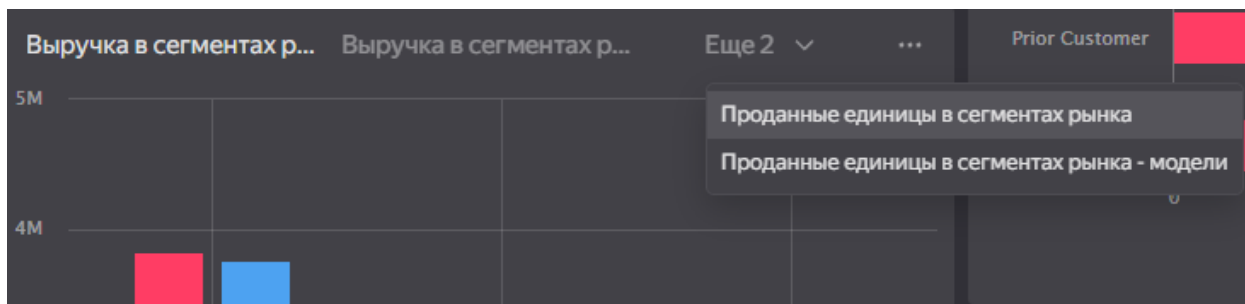


Рисунок 31. Выбор чартов на тему сегментации рынка



Рисунок 32. Выручка по сегментам и типам велосипедов



Рисунок 33. Выручка по сегментам и моделям велосипедов



Рисунок 34. Проданные единицы по сегментам и типам велосипедов



Рисунок 35. Проданные единицы по сегментам и моделям велосипедов

Следующая тема – географическое распределение продаж. Тут важно отметить, что данные по этому чарту следует смотреть при использовании селекторов по типу и моделям велосипедов.

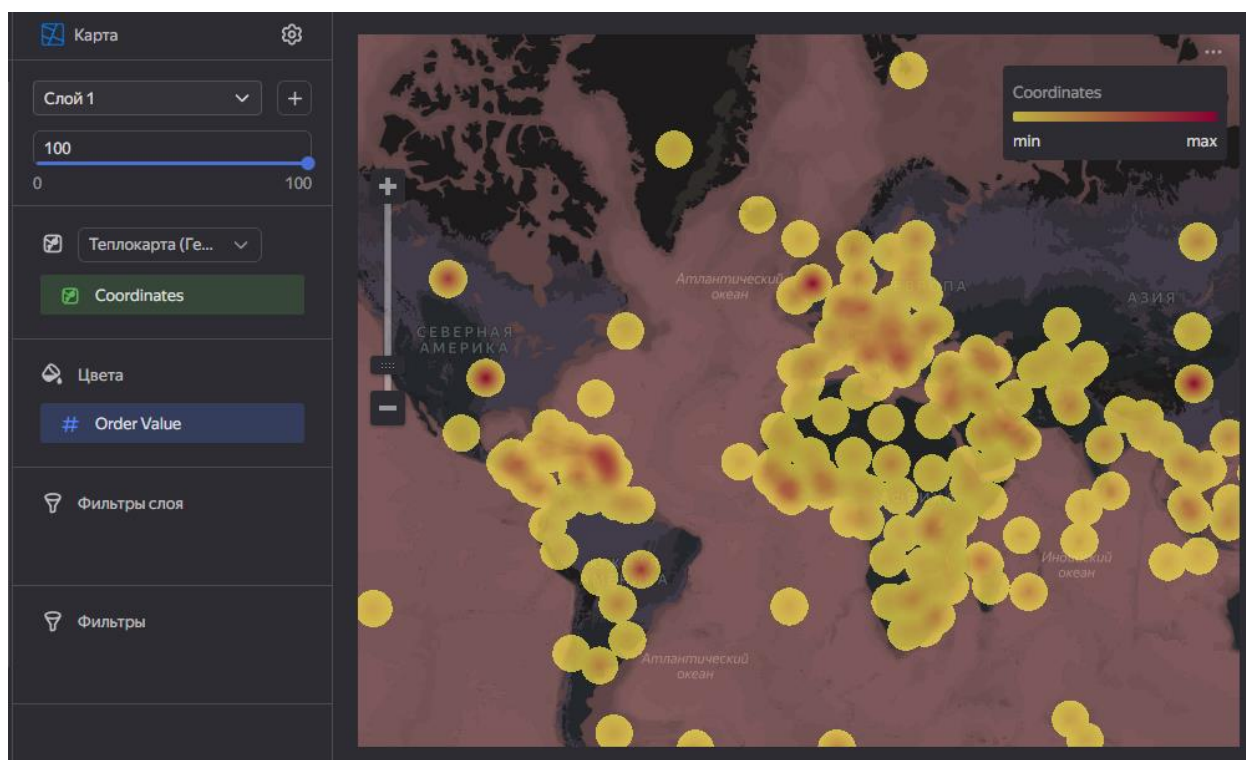


Рисунок 36. Тепловая карта

Далее: статус покупателя.

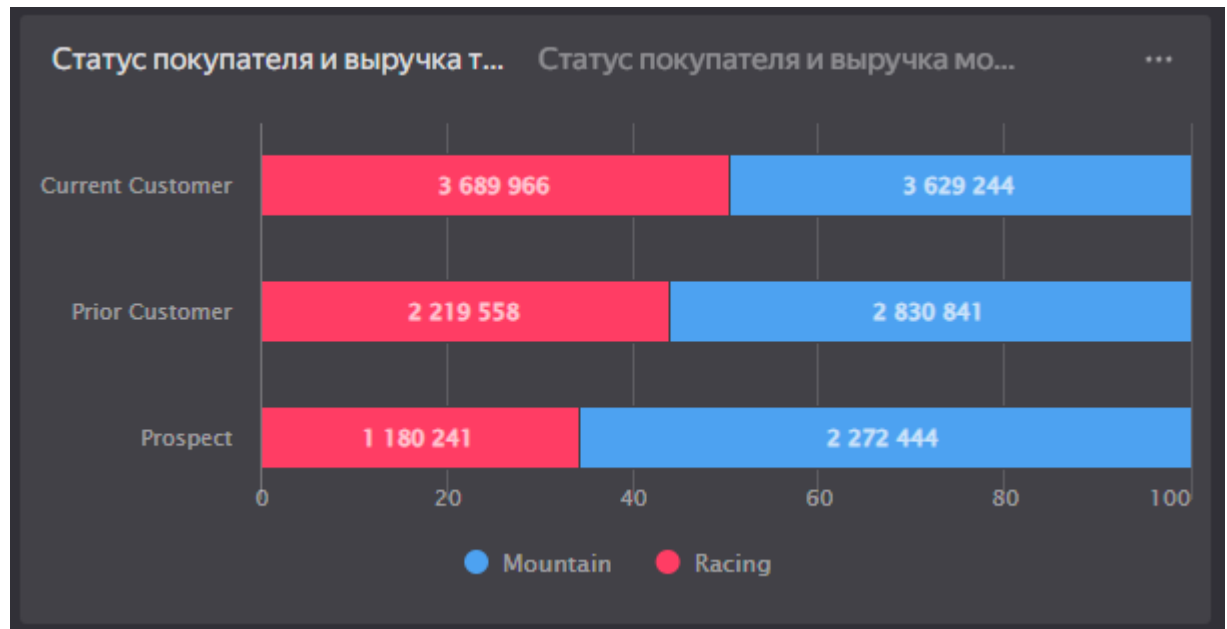


Рисунок 37. Выбор чарта

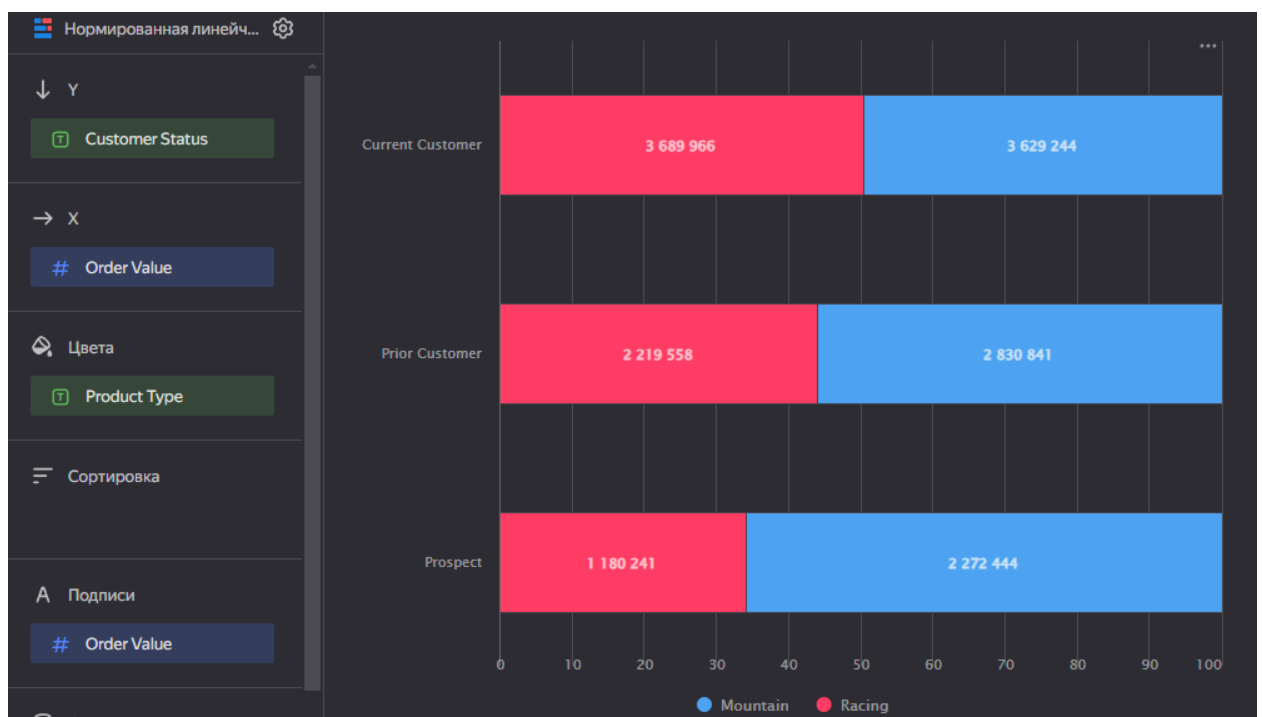


Рисунок 38. Статус покупателя и выручка типов велосипедов



Рисунок 39. Статус покупателя и выручка моделей

Последняя тематика чартов – динамика продаж категорий товаров.

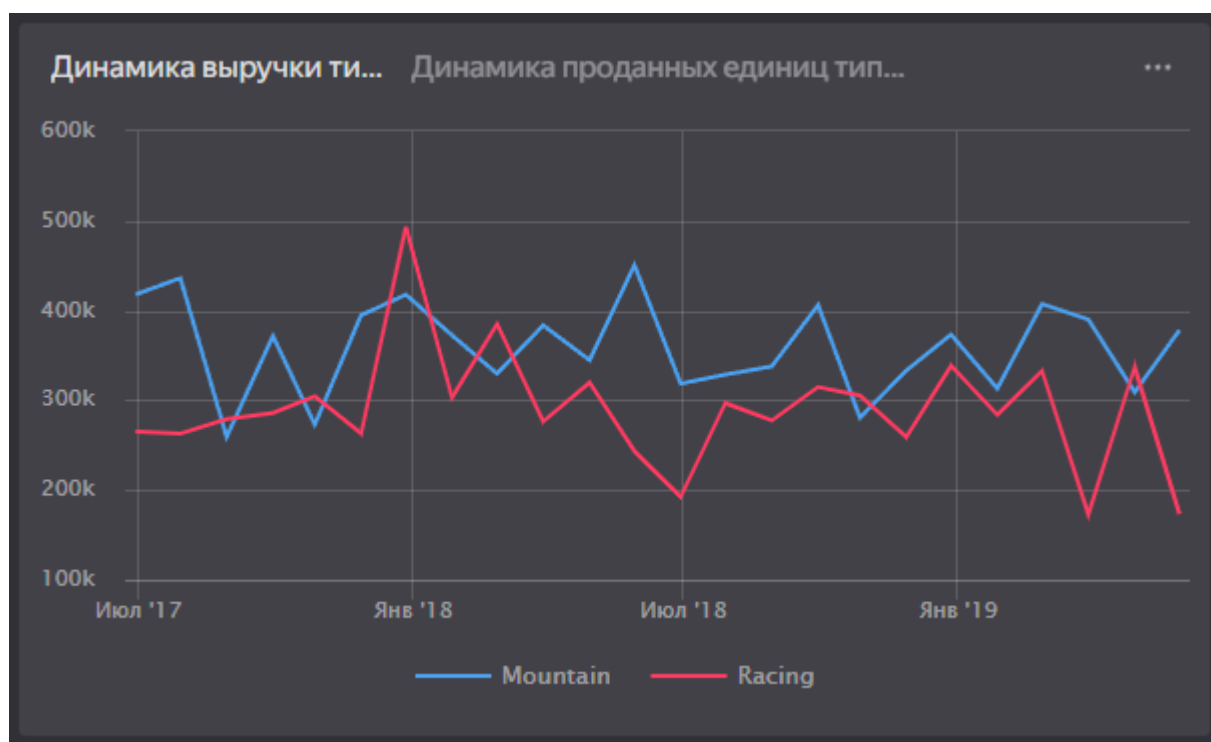


Рисунок 40. Выбор чартов

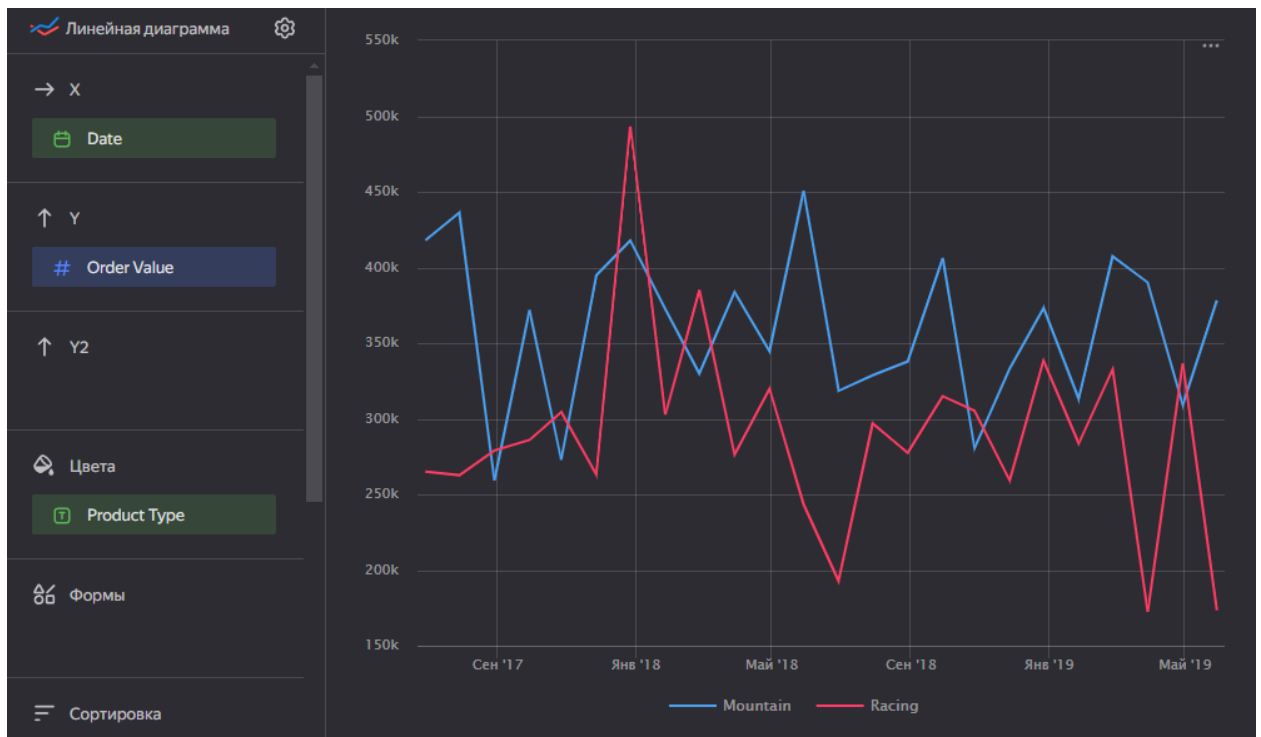


Рисунок 41. Динамика выручки типов

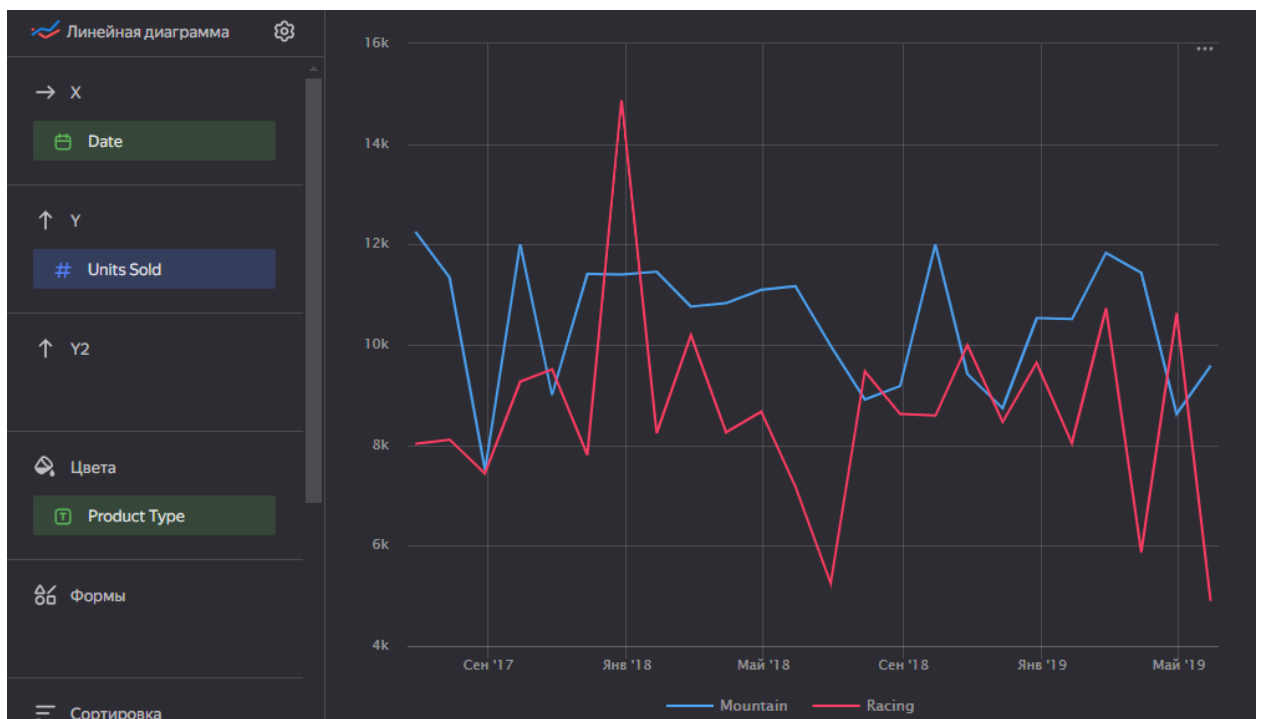


Рисунок 42. Динамика проданных единиц типов

Ссылка на дашборд: <https://datalens.yandex/n7y74g3jhmh6b>

Ссылка на Яндекс.Диск с файлами Excel: <https://disk.yandex.ru/d/IDmFMHMubJdLUA>

Краткие выводы

Из проведенного анализа видно, что категория велосипедов для гор является самой популярной среди потребителей, приносящей наибольшую выручку. Однако, в топ-3 по выручке входят две модели гоночных велосипедов: R100 BIKE и C990 Bike. Оптовая торговля гоночными велосипедами для гонок приносит наибольшую выручку, особенно модель R300 Bike.

Пик продаж гоночных велосипедов пришелся на декабрь 2017 года, являясь высшей точкой для всех категорий велосипедов с 2017 по 2019 годы. Спортивные (гоночные) велосипеды вызывают наибольший интерес у покупателей статуса Current, в то время как покупатели статуса Prospect проявляют к ним меньший интерес.

Также стоит отметить, что по географической карте можно увидеть, что в некоторых регионах предпочтение отдается только одной категории велосипедов, например, скоростным.

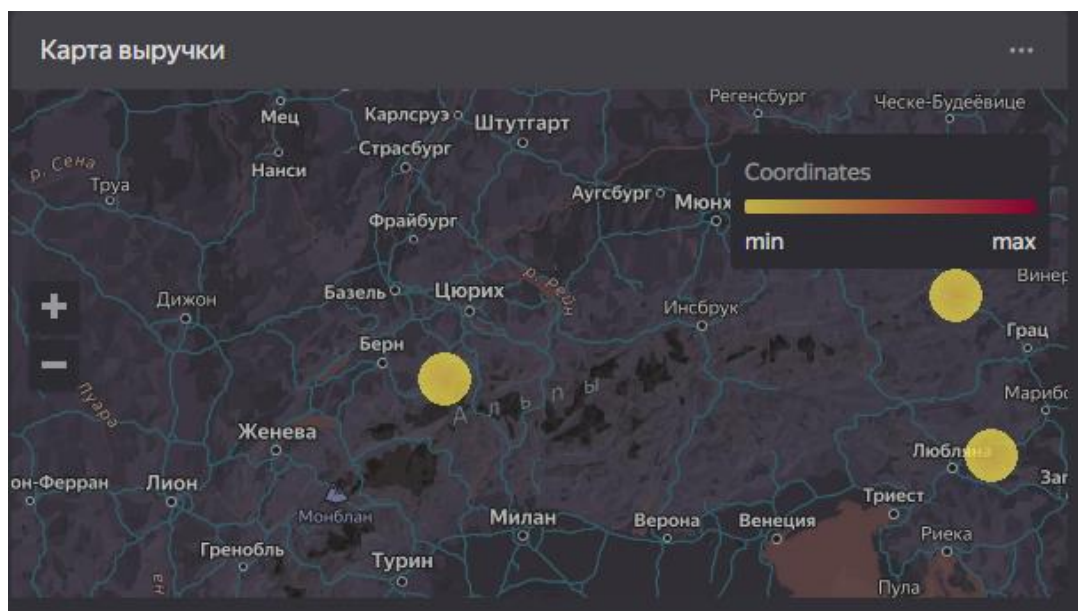


Рисунок 43. Карта с селектором Mountain

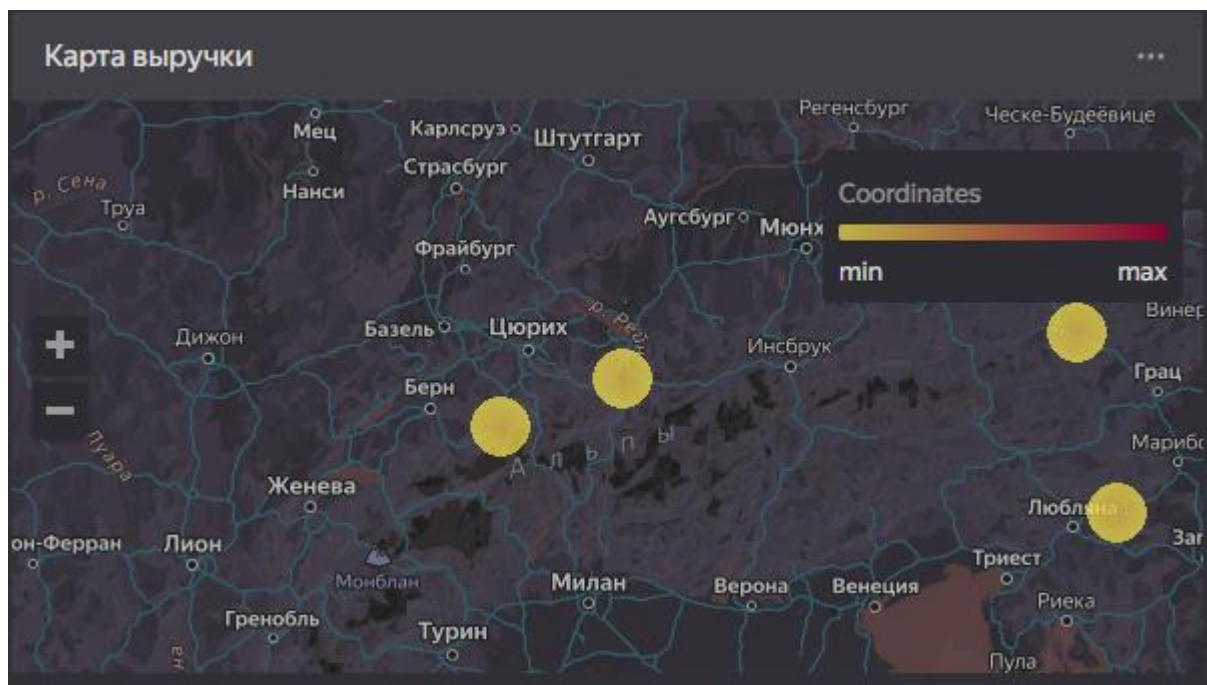


Рисунок 44. Карта с селектором типа Racing