



Московский институт электроники и
математики им. А.Н. Тихонова

Кафедра информационной
безопасности киберфизических
систем

Москва 2025

Расчет географических расстояний

Расчет географических расстояний



Использование инструментов RapidMiner для расчёта расстояний между географическими координатами. Задача — освоить процесс вычисления расстояний и познакомиться с основными геоинформационными формулами на примере реальных маршрутов.



Географические координаты



Географические координаты — способ указания местоположения любой точки Земли. Они включают широту (latitude, север-юг) и долготу (longitude, восток-запад). Значения указываются в градусах и десятичных долях.





Форматы координат

- Десятичные градусы (55.7558°): наиболее удобны для вычислений.
- Градусы, минуты, секунды ($55^\circ 45' 21''$): часто используются в картографии и навигации, требуют предварительной конвертации.

Исходные данные для расчёта

В расчёте используется маршрут Москва – Владивосток. Данные содержат города, координаты начальных и конечных точек сегментов маршрута. Всего 6 сегментов с координатами широты и долготы.



Import Data - Specify your data format

Specify your data format

☒ Header Row

Start Row

Column Separator

File Encoding

Escape Character

Decimal Character

☒ Use Quotes

☒ Skip Comments

☐ Trim Lines ☐ Multiline Text

	StartCity	StartLat	StartLon	EndCity	EndLat	EndLon
2	Moscow	55.7558	37.6176	Nizhny Novgorod	56.2965	43.9361
3	Nizhny Novgorod	56.2965	43.9361	Yekaterinburg	56.8389	60.6057
4	Yekaterinburg	56.8389	60.6057	Novosibirsk	55.0084	82.9357
5	Novosibirsk	55.0084	82.9357	Irkutsk	52.2869	104.3050
6	Irkutsk	52.2869	104.3050	Khabarovsk	48.4827	135.0840
7	Khabarovsk	48.4827	135.0840	Vladivostok	43.1155	131.8855



<new process*> – Altair AI Studio Educational 2024.1.0 @ MacBook-Air-Nasta-3.local

Views: Design Results Turbo Prep Auto Model Interactive Analysis

Process

Process

Read CSV

Parameters

Read CSV

Import CSV

column separators

trim lines

multiline text

use quotes

quotes character

	StartCity <i>polynomial</i>	StartLat <i>real</i>	StartLon <i>real</i>	EndCity	EndLat	EndLon <i>real</i>
1	Moscow	55.756	37.618		297	43.936
2	Nizhny Novgorod	56.297	43.936		839	60.606
3	Yekaterinburg	56.839	60.606		008	82.936
4	Novosibirsk	55.008	82.936	Irkutsk	287	104.305
5	Irkutsk	52.287	104.305	Khabarovsk	483	135.084
6	Khabarovsk	48.483	135.084	Vladivostok	115	131.886

Change Type

- polynomial
- binominal
- real
- integer
- date_time
- date
- time

Импорт и проверка данных

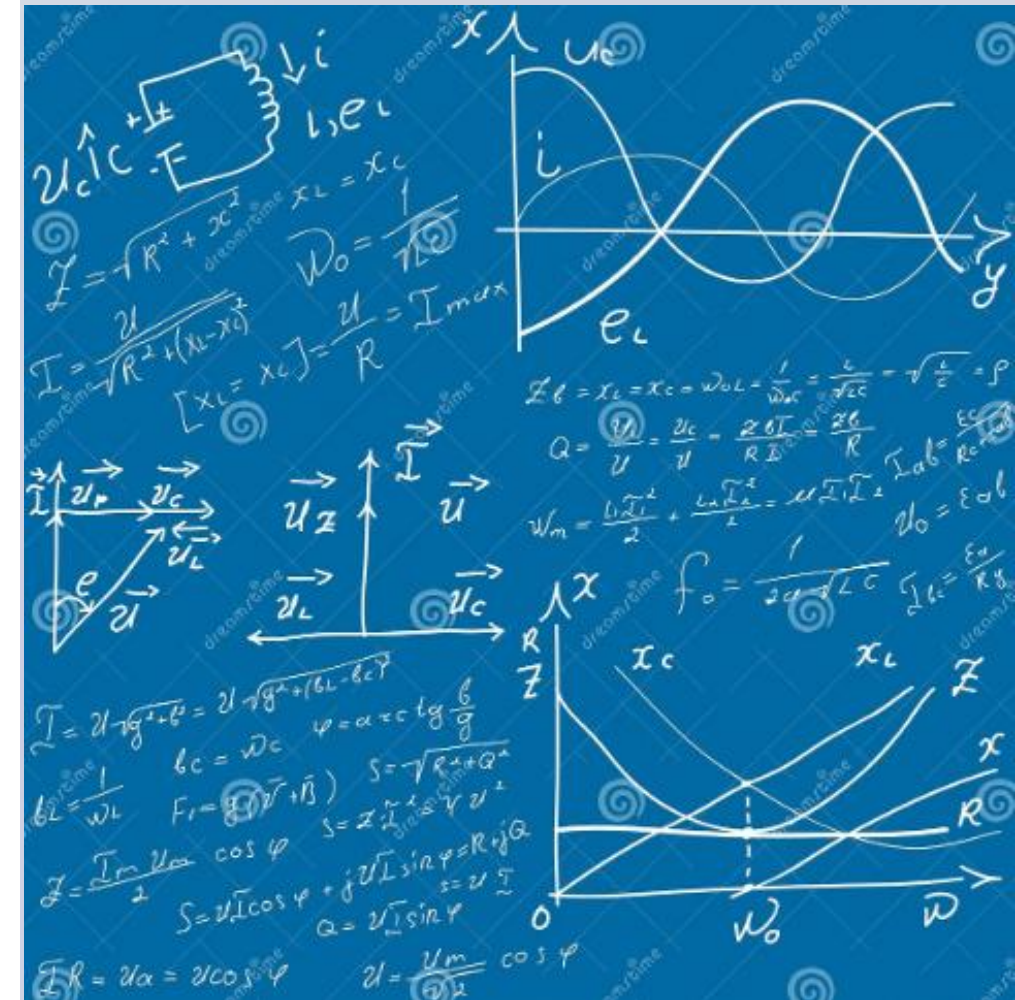
Данные загружаются оператором Read CSV. Проверяется правильность формата и тип столбцов (широта, долгота — тип real). Ошибки на этом этапе критически влияют на дальнейшие расчёты.

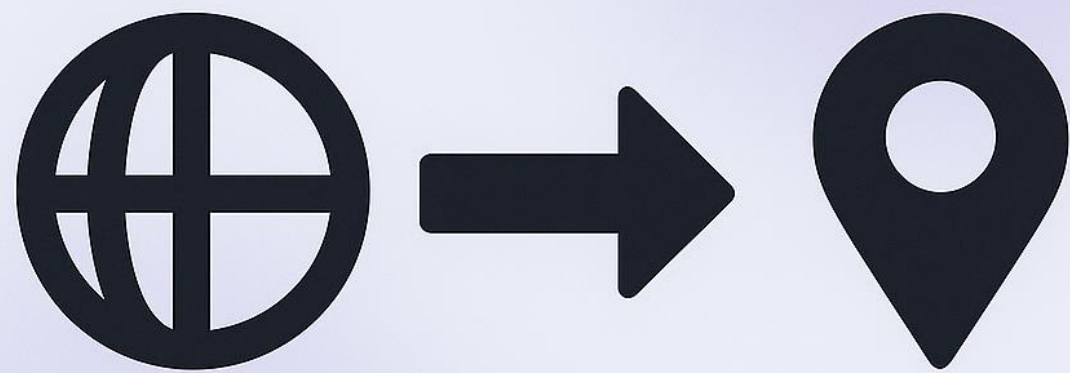
Преобразование координат

Для тригонометрических расчётов координаты из градусов переводят в радианы. Используется формула:

$$\text{Радианы} = \text{Градусы} \times \pi / 180$$

Это обязательный шаг перед расчётом расстояний на сфере.





Формула расчёта расстояния

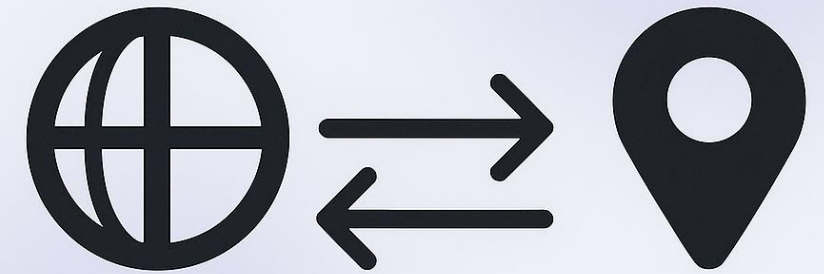
Расстояние вычисляется по формуле сферического закона косинусов (Haversine). Это точная тригонометрическая формула, учитывающая сферическую форму Земли и обеспечивающая высокую точность на практике.

Формула Haversine



$$\text{distance_km} = 6371 \times \arccos[\sin(\text{lat1}) \times \sin(\text{lat2}) + \cos(\text{lat1}) \times \cos(\text{lat2}) \times \cos(\text{lon2} - \text{lon1})]$$

где 6371 — средний радиус Земли в километрах.



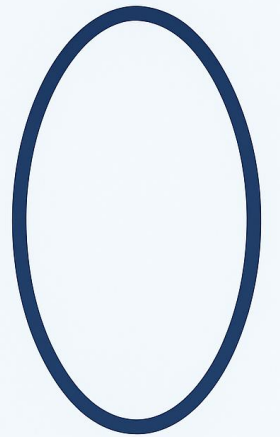


Почему сферическая модель

Формула Haversine предполагает Землю идеальной сферой. На практике её погрешность менее 1% на дистанциях до 5000 км, что приемлемо для большинства задач логистики и планирования маршрутов.

Эллипсоидные модели Земли

Для более точных задач (геодезия, авиация) используют формулы на основе эллипсоида (например, формула Винценти). Точность таких формул доходит до сантиметров даже на больших дистанциях.





Edit Parameter List: function descriptions
List of functions to generate.

column name	function expressions
lat1_rad	StartLat * 3.1416 / 180
lon1_rad	StartLon * 3.1416 / 180
lat2_rad	EndLat * 3.1416 / 180
lon2_rad	EndLon * 3.1416 / 180

Add Entry

Remove Entry

Apply

Cancel

Edit Expression: function expressions

Expression

1 6371 * acos(sin(lat1_rad) * sin(lat2_rad) + cos(lat1_rad) * cos(lat2_rad) * cos(lon2_rad - lon1_rad))

Info: Expression is syntactically correct.

Functions

Logical

Comparison

Text information

Text transformation

Mathematical functions

Statistical functions

Trigonometric functions

Rounding functions

Conversion functions

Date calculation

Time calculation

Inputs

Regular Attributes

EndCity

EndLat

EndLon

StartCity

StartLat

StartLon

Basic Constants

Date Function Constants

Macros

Apply

Cancel

Оператор Generate Attributes

Этот оператор используется для расчёта новых столбцов в RapidMiner. Сначала переводятся градусы в радианы, затем вычисляются расстояния. Каждый шаг прописывается последовательно в виде формул.

Типичные ошибки при расчёте

- Неправильно указаны знаки координат.
- Координаты перепутаны (широта-долгота).
- Использован неверный формат градусов.



Такие ошибки значительно искажают итоговые результаты.



Контроль корректности расчёта

Результаты проверяются
сравнением с известными
расстояниями (например,
Москва–Нижний Новгород ~400
км). Совпадение с реальными
расстояниями подтверждает
правильность расчёта и
корректность данных.

ExampleSet (Generate Attributes) X ExampleSet (/ / Temporary Repository / route - Лист1) X											
Open in Turbo Prep Auto Model Interactive Analysis Filter (6 / 6 examples): all											
Row No.	StartCity	StartLat	StartLon	EndCity	EndLat	EndLon	lat1_rad	lon1_rad	lat2_rad	lon2_rad	distance_km
1	Moscow	55.756	37.618	Nizhny Novg...	56.297	43.936	0.973	0.657	0.983	0.767	397.045
2	Nizhny Novg...	56.297	43.936	Yekaterinburg	56.839	60.606	0.983	0.767	0.992	1.058	1020.469
3	Yekaterinburg	56.839	60.606	Novosibirsk	55.008	82.936	0.992	1.058	0.960	1.448	1399.578
4	Novosibirsk	55.008	82.936	Irkutsk	52.287	104.305	0.960	1.448	0.913	1.820	1434.565
5	Irkutsk	52.287	104.305	Khabarovsk	48.483	135.084	0.913	1.820	0.846	2.358	2205.331
6	Khabarovsk	48.483	135.084	Vladivostok	43.115	131.886	0.846	2.358	0.753	2.302	646.115

Расчёт суммарной длины маршрута

Используя оператор Aggregate, вычисляется общая длина маршрута путём суммирования отдельных сегментов. Например, суммарное расстояние от Москвы до Владивостока ≈ 7100 км.

Edit Parameter List: aggregation attributes
The attributes which should be aggregated.

aggregation attribute	aggregation functions
distance_km	sum
distance_km	maximum

Add Entry Remove Entry Apply Cancel

Result History ExampleSet (Aggregate)

Open in Turbo Prep Auto Model

Row No.	sum(distan...	maximum(...
1	7103.103	2205.331

Data Statistics





Определение экстремумов маршрута

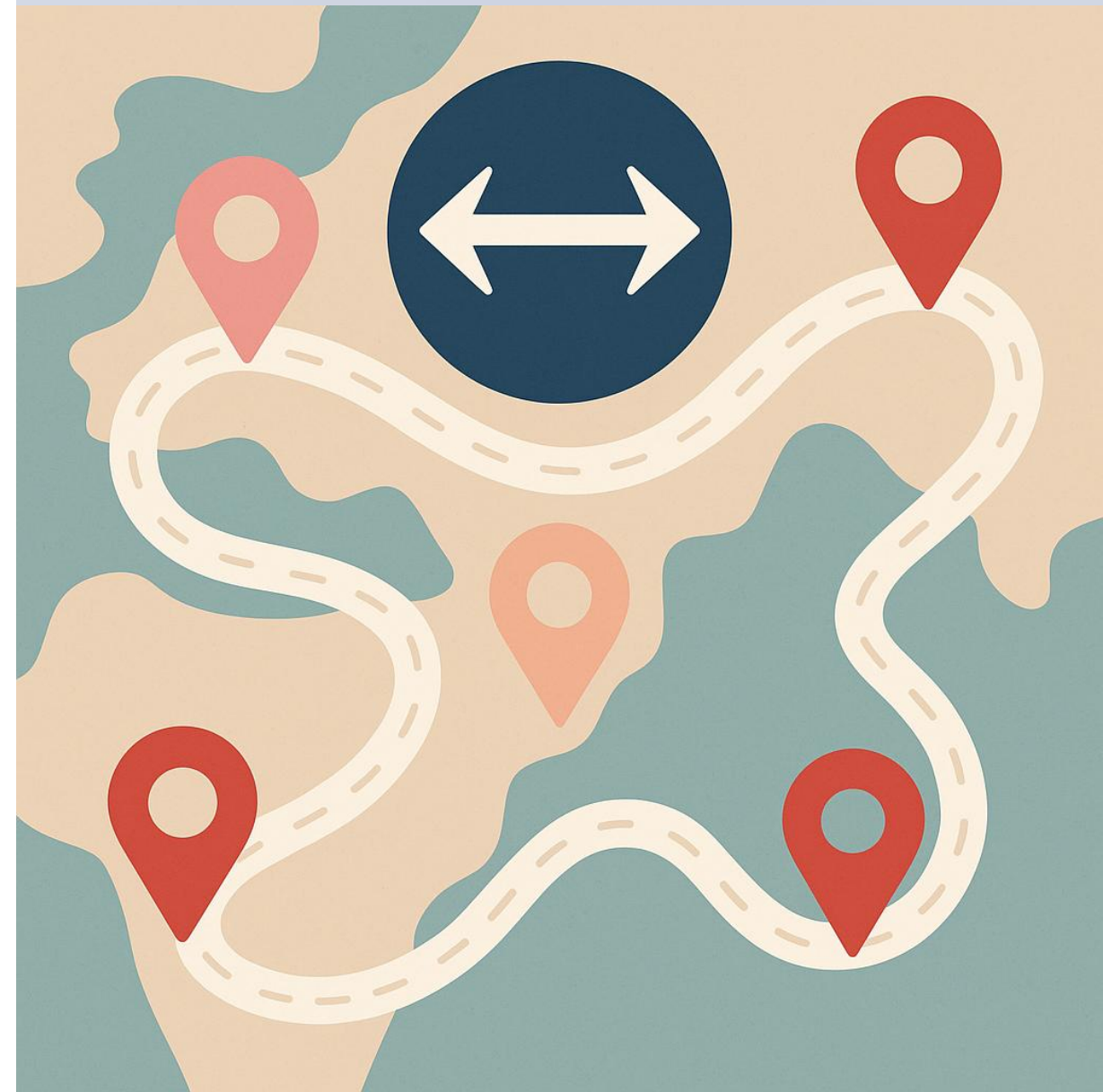
Дополнительно вычисляются минимальный и максимальный сегменты маршрута. Например, максимальный сегмент маршрута Иркутск–Хабаровск (~2200 км) выделяется как наиболее протяжённый участок.

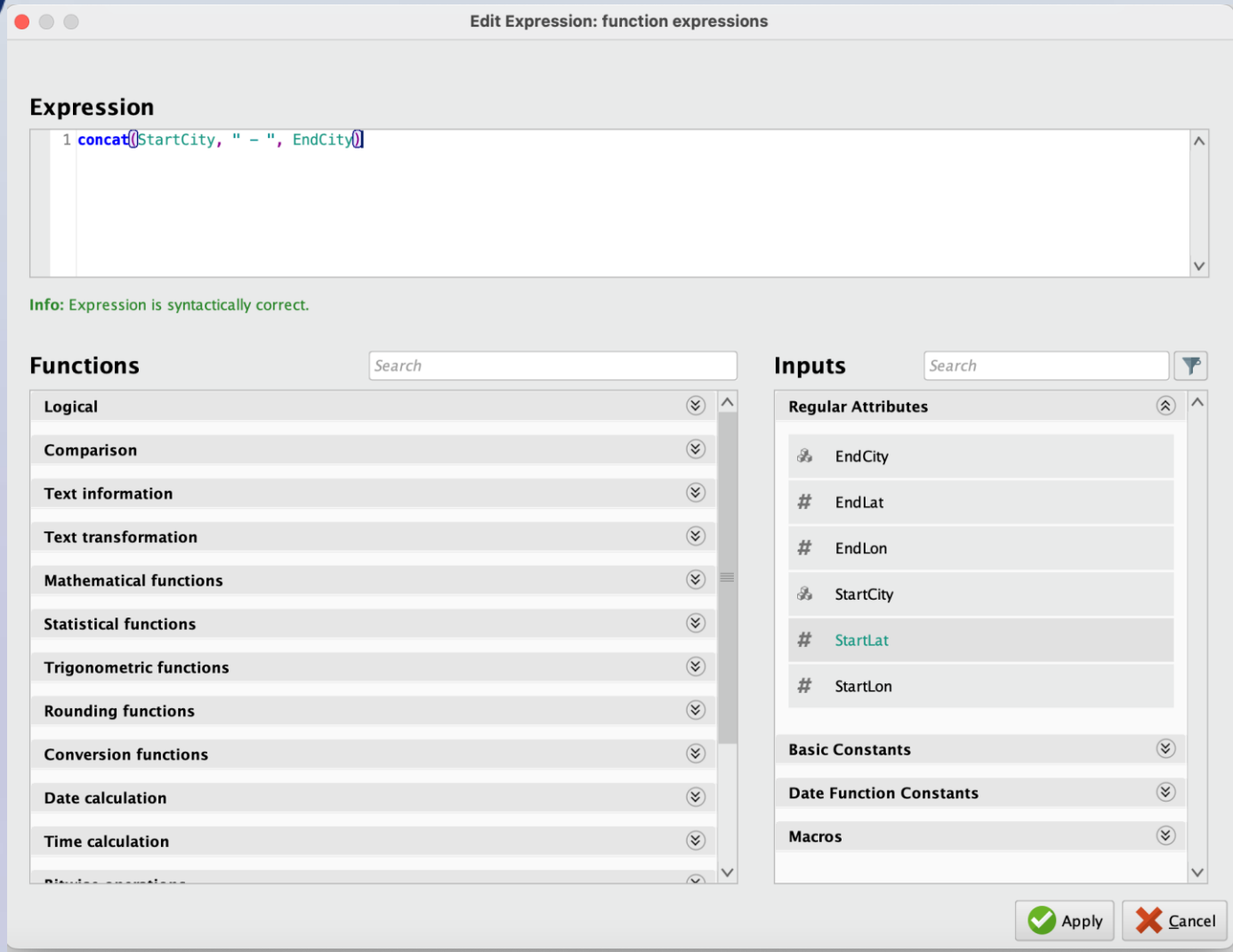
Open in Turbo Prep Auto Model Interactive Analysis Filter (6 / 6 examples): all

Row No.	StartCity	StartLat	StartLon	EndCity	EndLat	EndLon	lat1_rad	lon1_rad	lat2_rad	lon2_rad	distanc... ↓
5	Irkutsk	52.287	104.305	Khabarovsk	48.483	135.084	0.913	1.820	0.846	2.358	2205.331
4	Novosibirsk	55.008	82.936	Irkutsk	52.287	104.305	0.960	1.448	0.913	1.820	1434.565
3	Yekaterinburg	56.839	60.606	Novosibirsk	55.008	82.936	0.992	1.058	0.960	1.448	1399.578
2	Nizhny Novg...	56.297	43.936	Yekaterinburg	56.839	60.606	0.983	0.767	0.992	1.058	1020.469
6	Khabarovsk	48.483	135.084	Vladivostok	43.115	131.886	0.846	2.358	0.753	2.302	646.115
1	Moscow	55.756	37.618	Nizhny Novg...	56.297	43.936	0.973	0.657	0.983	0.767	397.045

Расчёт среднего расстояния

Среднее расстояние сегментов маршрута позволяет оценить его равномерность. Большая вариация расстояний показывает необходимость учёта логистических особенностей при планировании остановок и снабжения.





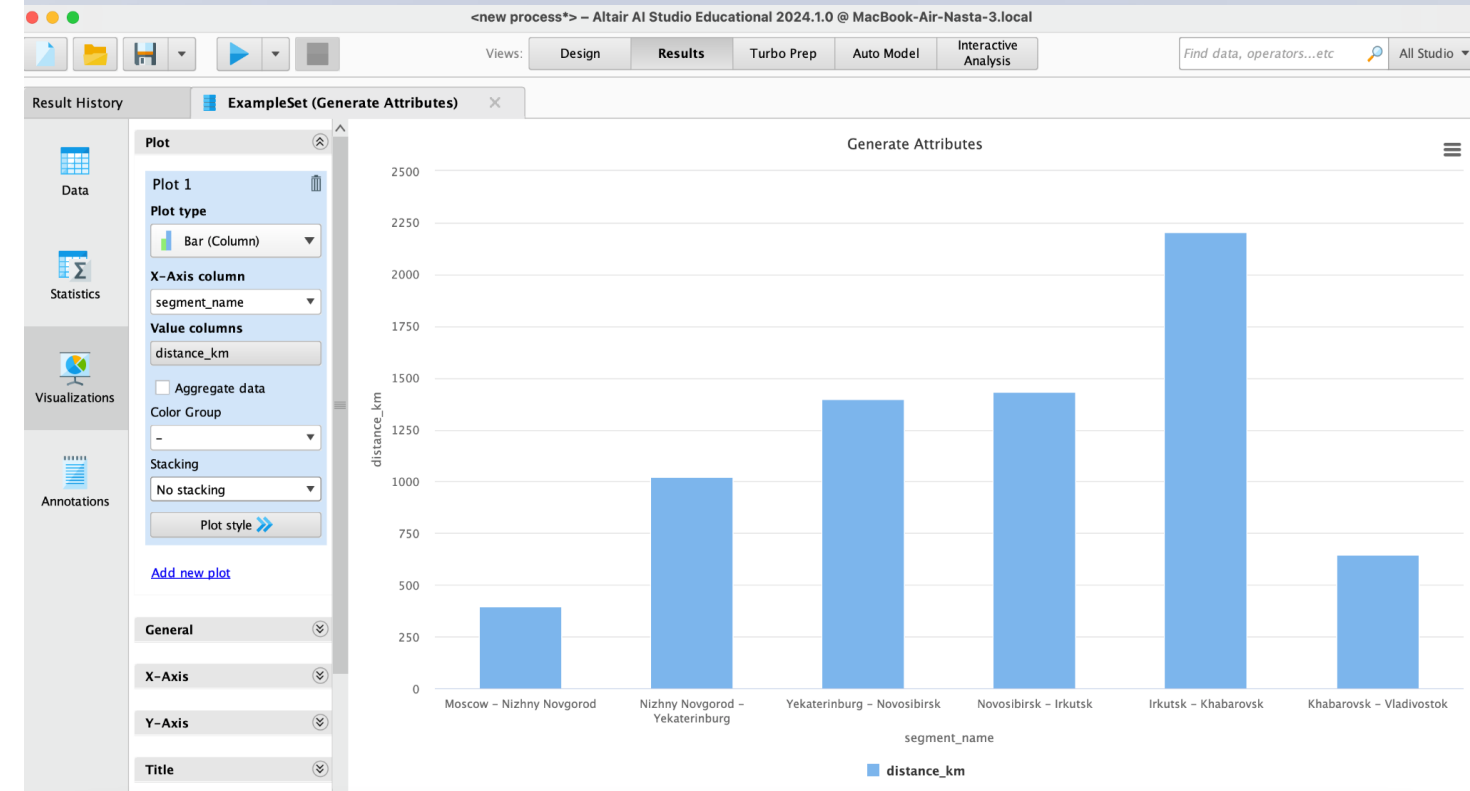
Row No.	StartCity	StartLat	StartLon	EndCity	EndLat	EndLon	lat1_rad	lon1_rad	lat2_rad	lon2_rad	distance_km	segment_n...
1	Moscow	55.756	37.618	Nizhny Novg...	56.297	43.936	0.973	0.657	0.983	0.767	397.045	Moscow - Ni...
2	Nizhny Novg...	56.297	43.936	Yekaterinburg	56.839	60.606	0.983	0.767	0.992	1.058	1020.469	Nizhny Novg...
3	Yekaterinburg	56.839	60.606	Novosibirsk	55.008	82.936	0.992	1.058	0.960	1.448	1399.578	Yekaterinbu...
4	Novosibirsk	55.008	82.936	Irkutsk	52.287	104.305	0.960	1.448	0.913	1.820	1434.565	Novosibirsk ...
5	Irkutsk	52.287	104.305	Khabarovsk	48.483	135.084	0.913	1.820	0.846	2.358	2205.331	Irkutsk - Kh...
6	Khabarovsk	48.483	135.084	Vladivostok	43.115	131.886	0.846	2.358	0.753	2.302	646.115	Khabarovsk ...

Создание названий сегментов

В RapidMiner удобно использовать оператор Generate Attributes для формирования столбца с названиями сегментов маршрута, объединяя начальные и конечные города (например, «Москва – Нижний Новгород»).

Визуализация данных

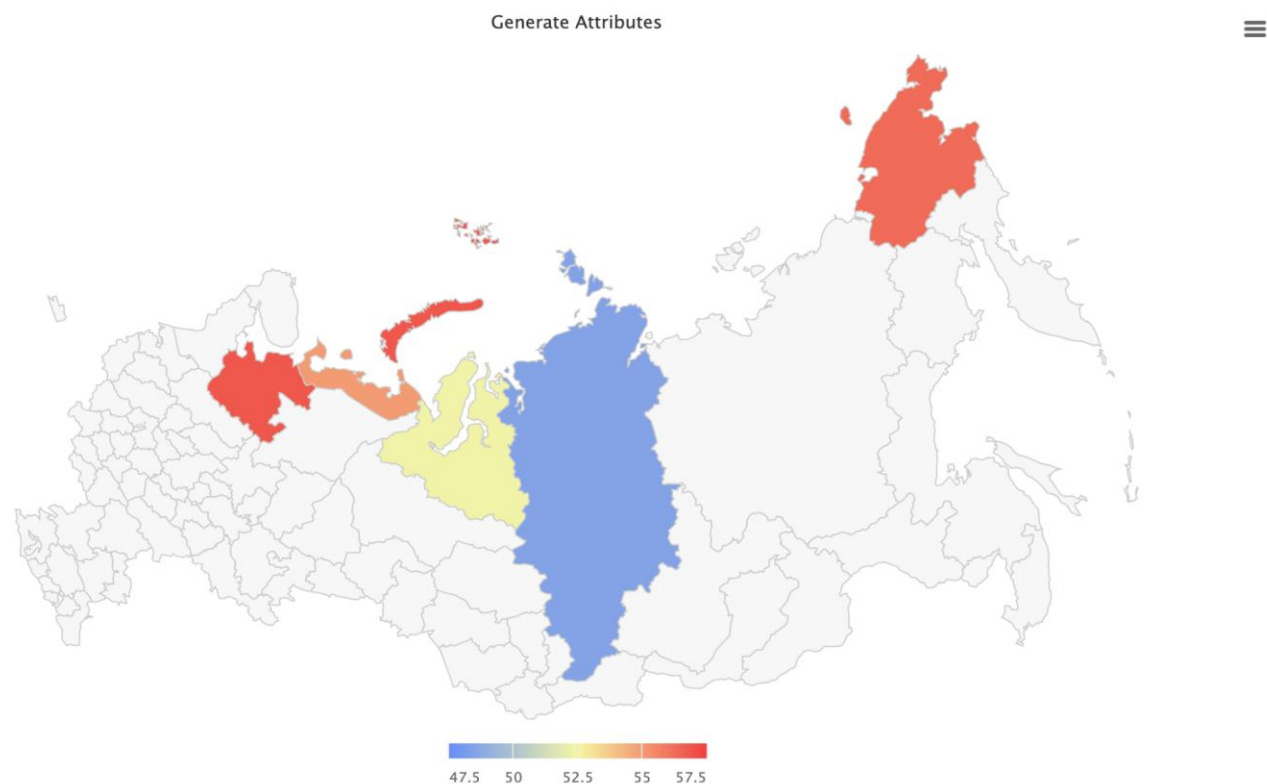
Полученные расстояния сегментов удобно визуализировать на бар-чарте, где ось X — название сегмента, а ось Y — дистанция. Это позволяет быстро выявить наиболее длинные участки маршрута.





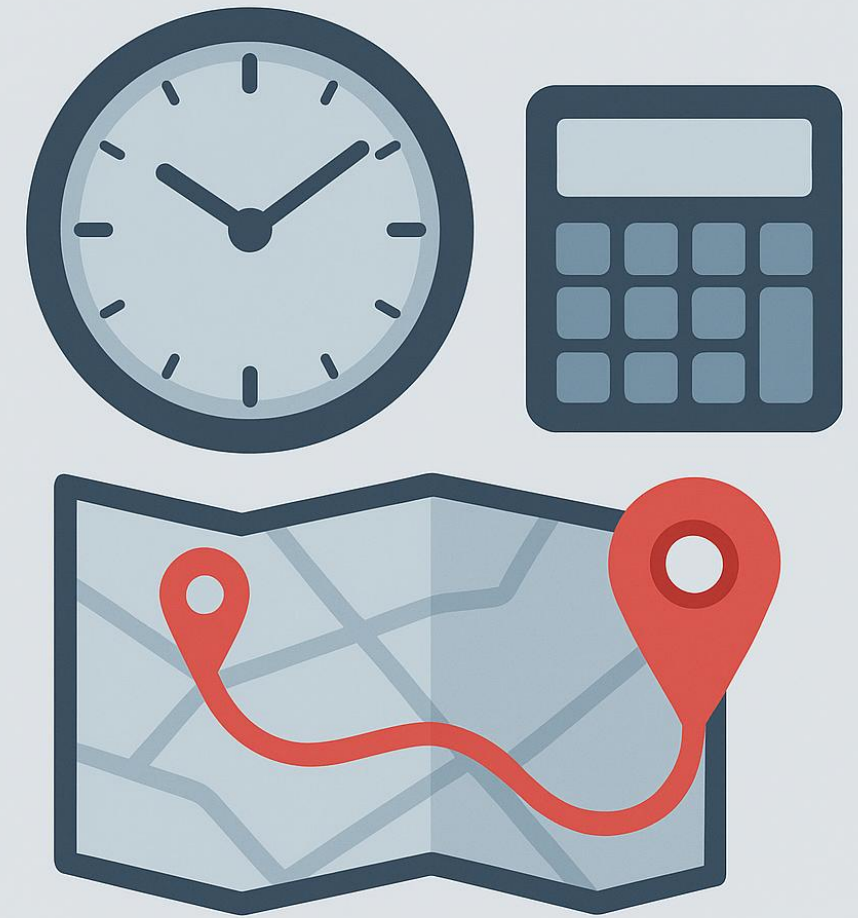
Картографическая визуализация

Для наглядности результаты расчётов можно визуализировать на карте (choropleth map), показывая цветом длину сегментов маршрута. Такие карты упрощают восприятие и интерпретацию данных.



Дополнительные расчёты (время)

Добавив среднюю скорость транспорта, можно рассчитать продолжительность каждого сегмента маршрута. Это важный элемент планирования расписаний и оптимизации логистических процессов.



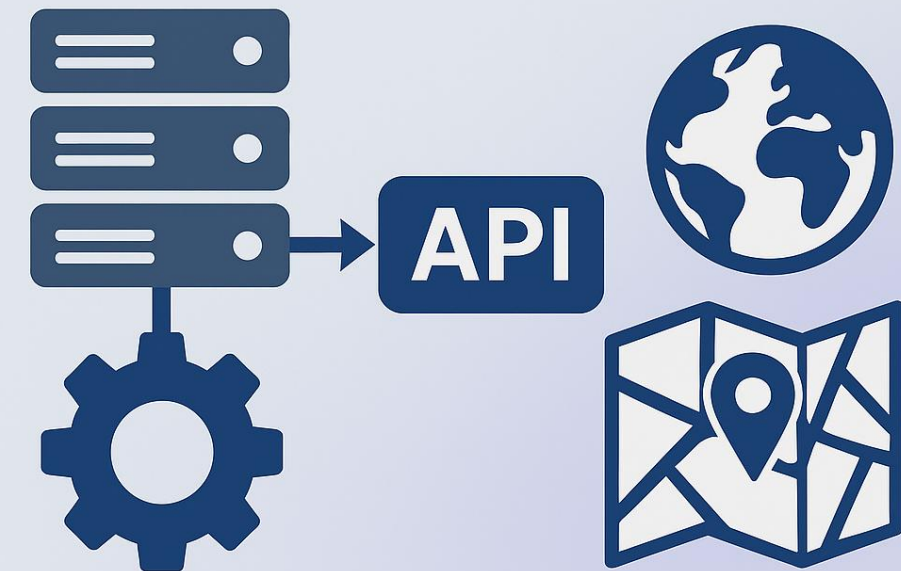


Практическое применение

Географические расчёты востребованы в бизнесе для оптимизации транспортных затрат, выбора мест расположения складов и пунктов выдачи товаров, а также в туристической отрасли.

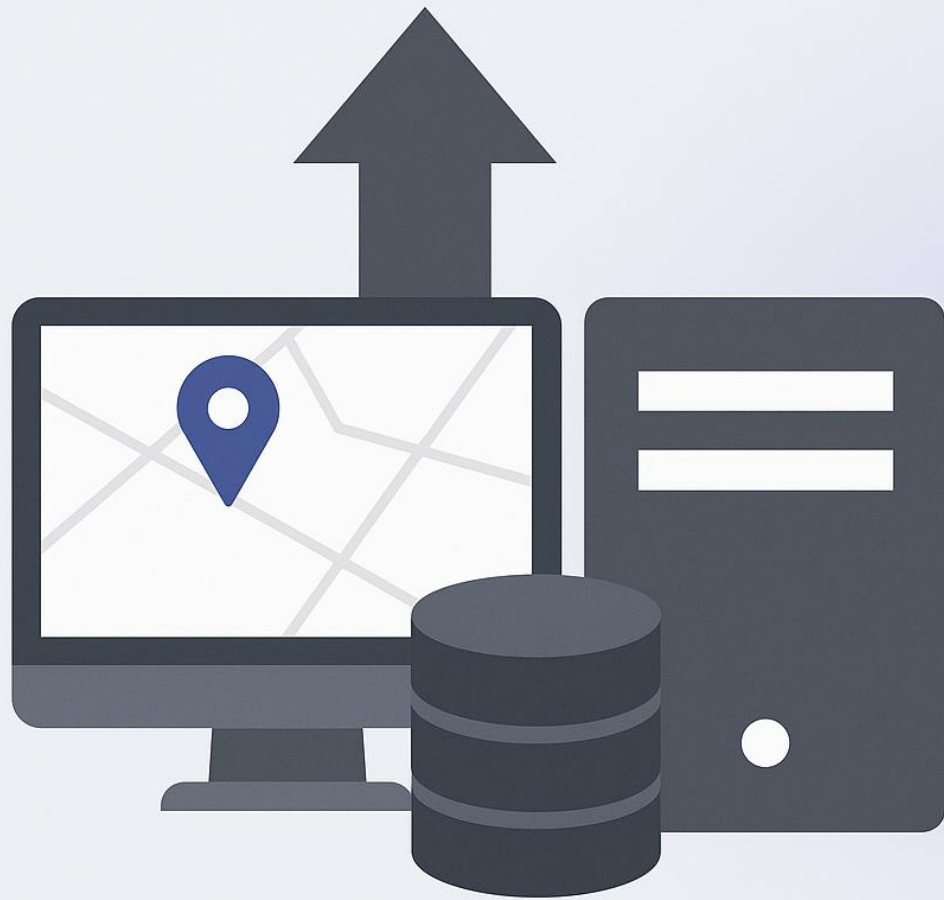
Интеграция с внешними источниками

RapidMiner позволяет интегрироваться с внешними API, такими как Google Maps или OpenStreetMap, загружая данные о координатах и автоматизируя их расчёт в реальном времени.





Масштабирование процессов



Расчёты в RapidMiner легко масштабируются до тысяч или миллионов пар координат. Для больших объёмов данных рекомендуется использовать пакетную обработку или серверные расширения платформы.

Выводы и рекомендации

RapidMiner обеспечивает точный расчёт географических расстояний с минимальными временными затратами.

Использование визуализации и дополнительных аналитических функций повышает качество и практическую ценность результатов.

