

Automated issues - visualisations

```

    ▶ Run   Limit 100   Format   View history
1 select distinct created_at, lat, lon, item_id, date(reviewed_at) as dt_review, current_tags
2   , CASE WHEN contains(current_tags,'auto:triage:home-trip')=true THEN 'auto_triage_home_trip'
3     WHEN contains(current_tags,'auto:triage:destination-place')=true THEN 'auto_triage_destination_place'
4     else 'valid_trip' end as trip_duration,
5
6   CASE WHEN contains(current_tags,'possible_bad_arrival_RP_distance')=true THEN 'RP_distance>50m'
7     else null end as RP_distance_auto,
8
9   CASE WHEN contains(current_tags,'possible_bad_arrival_backdoor')=true THEN 'possible_bad_backdoor'
10  else null end as arrival_backdoor_auto,
11
12  case when destination_search_result_administrative_unit_type like 'neighborhood%' then 'neighborhood'
13  when destination_search_result_administrative_unit_type like 'country' then 'country'
14  when destination_search_result_administrative_unit_type like 'region'
15  when destination_search_result_administrative_unit_type like 'place' then 'place'
16  when destination_search_result_administrative_unit_type like 'street%' then 'street'
17  when destination_search_result_administrative_unit_type like 'poi%' then 'pol'
18  when destination_search_result_administrative_unit_type like 'address' then 'address'
19  when destination_search_result_administrative_unit_type like 'district%' then 'district'
20  when destination_search_result_administrative_unit_type like 'postcode' then 'postcode'
21  when destination_search_result_administrative_unit_type like 'locality%' then 'locality'
22  when destination_search_result_administrative_unit_type is null then null
23  else 'other' end destination_search_unit_type
24
25
26  , case when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing')) and (contains(current_tags,'address'))
27  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:incorrect')) and (contains(current_tags,'POI')) then
28  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:incorrect_location')) and (contains(current_tags,'address'))
29  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing')) and (contains(current_tags,'POI')) then
30  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing')) and (contains(current_tags,'not_exist')) and (contains(current_tags,'address')) then
31  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing_meta_data')) and (contains(current_tags,'address'))
32  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing_meta_data')) and (contains(current_tags,'POI')) then
33  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing_meta_data')) and (contains(current_tags,'POI')) then
34  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing_meta_data')) and (contains(current_tags,'POI')) then
35  when current_status = 'open' and (contains(current_tags,'3:missing')) and (contains(current_tags,'address')) then 'missing_address' ● Ready
36  when current_status = 'open' and (contains(current_tags,'3:missing')) and (contains(current_tags,'POI')) then 'missing_POI'
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72

```

```

    ▶ Run   Limit 100   Format   View history
37 when current_status = 'open' and (contains(current_tags,'3:incorrect_location')) and (contains(current_tags,'address')) then 'incorrect_location_address'
38 when current_status = 'open' and (contains(current_tags,'3:incorrect_location')) and (contains(current_tags,'POI')) then 'incorrect_location_POI'
39 when current_status = 'open' and (contains(current_tags,'3:incorrect')) and (contains(current_tags,'address')) then 'incorrect_address'
40 when current_status = 'open' and (contains(current_tags,'3:missing')) and (contains(current_tags,'POI')) then 'not_exist_POI'
41 when current_status = 'open' and (contains(current_tags,'3:missing_meta_data')) and (contains(current_tags,'address')) then 'missing_meta_data_address'
42 when current_status = 'open' and (contains(current_tags,'3:missing_meta_data')) and (contains(current_tags,'POI')) then 'missing_meta_data_POI'
43 when current_status = 'open' and (contains(current_tags,'3:wrong_category')) and (contains(current_tags,'POI')) then 'wrong_category_POI'
44 when current_status = 'open' and (contains(current_tags,'3:wrong_category')) and (contains(current_tags,'undefined')) then 'undefined'
45 when current_status = 'open' and (contains(current_tags,'1:oneaway')) then 'oneaway'
46 when current_status = 'open' and (contains(current_tags,'1:turn-restriction')) then 'turn-restriction'
47 when current_status = 'open' and (contains(current_tags,'1:illegal_close')) then 'illegal_close'
48 when current_status = 'open' and (contains(current_tags,'1:missing')) then 'missing'
49 when current_status = 'open' and (contains(current_tags,'1:road_geometry')) then 'road_geometry'
50 when current_status = 'open' and (contains(current_tags,'1:road_connectivity')) then 'road_connectivity'
51 when current_status = 'open' and (contains(current_tags,'1:road_classification')) then 'road_classification'
52 when current_status = 'open' and (contains(current_tags,'1:vehicle_restriction')) then 'vehicle_restriction'
53 when current_status = 'open' and (contains(current_tags,'1:access')) then 'access'
54 when current_status = 'open' and (contains(linked_issues,'Jira/RTFDB-769'))
55 when current_status = 'open' and cardinality(filter(array_distinct(current_tags), x => x like '%2%')) > 0 and (contains(linked_issues,'Jira/RTFDB-769'))
56 when current_status = 'open' and cardinality(filter(array_distinct(current_tags), x => x like '%2%')) > 0 and contains(linked_issues,'Jira/RTCO-3088')
57 when current_status = 'open' and cardinality(filter(array_distinct(current_tags), x => x like '%2%')) > 0 and contains(linked_issues,'Jira/RTCO-4765')
58 when current_status = 'open' and cardinality(filter(array_distinct(current_tags), x => x like '%2%')) > 0 and contains(linked_issues,'Jira/RTCO-447')
59 when current_status = 'open' and cardinality(filter(array_distinct(current_tags), x => x like '%2%')) > 0 and contains(linked_issues,'Jira/RTCO-449')
60 when current_status = 'open' and cardinality(filter(array_distinct(current_tags), x => x like '%2%')) > 0 and contains(linked_issues,'Jira/RTCO-9523')
61 when current_status = 'open' and cardinality(filter(array_distinct(current_tags), x => x like '%2%')) > 0 and contains(linked_issues,'Jira/RTCO-9784')
62 when current_status = 'open' and cardinality(filter(array_distinct(current_tags), x => x like '%2%')) > 0 and contains(linked_issues,'Jira/NA-1064')
63 else 'other' end as reason
64
65 , case when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing')) and (contains(current_tags,'address'))
66  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing')) and (contains(current_tags,'POI')) then
67  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:incorrect_location')) and (contains(current_tags,'address'))
68  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:incorrect')) and (contains(current_tags,'POI')) then
69  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing')) and (contains(current_tags,'not_exist')) and (contains(current_tags,'address')) then
70  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing')) and (contains(current_tags,'not_exist')) and (contains(current_tags,'POI')) then
71  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing_meta_data')) and (contains(current_tags,'POI')) then
72  when current_status = 'open' and (contains(current_tags,'RP')) and (contains(current_tags,'3:missing_meta_data')) and (contains(current_tags,'POI')) then

```

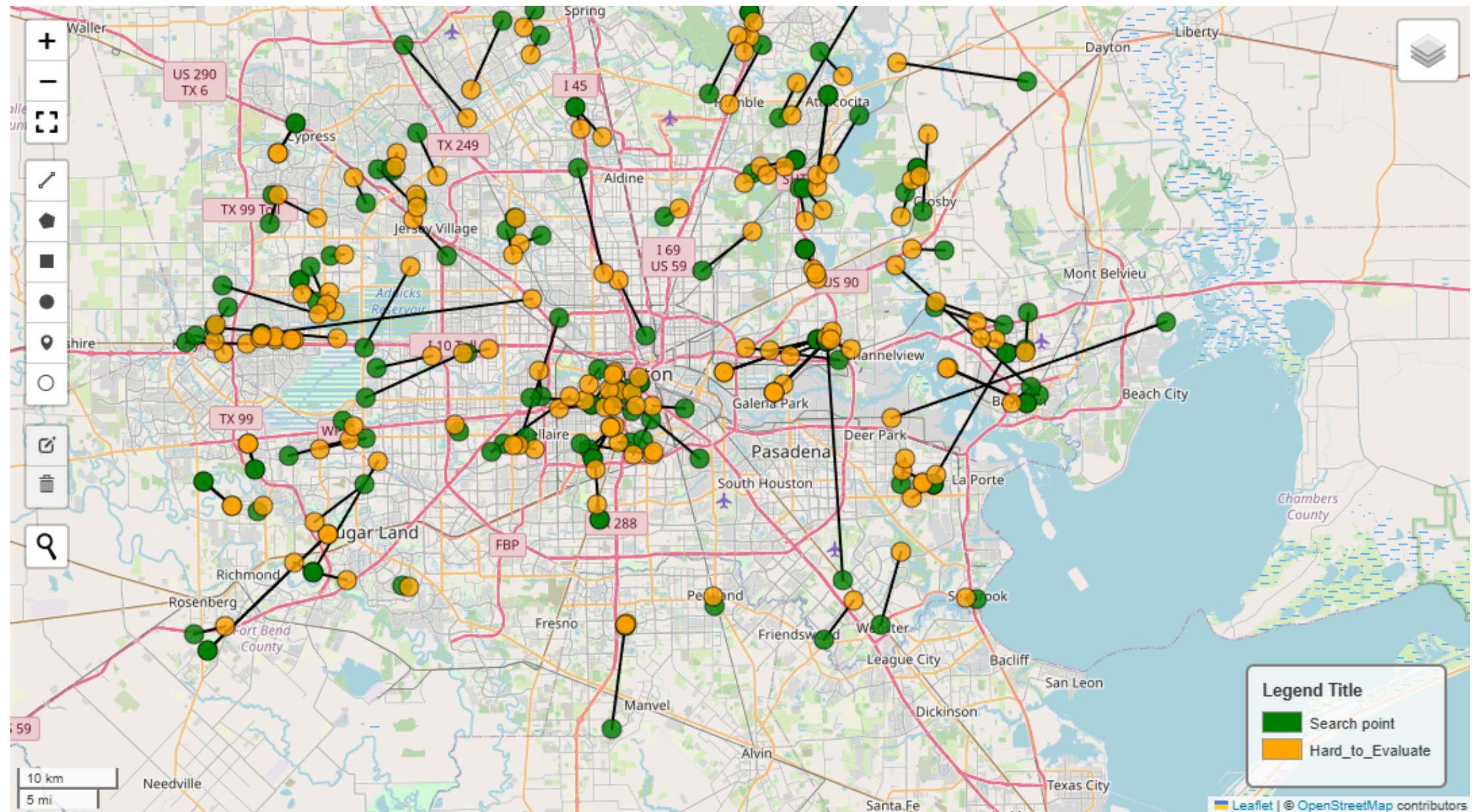
```

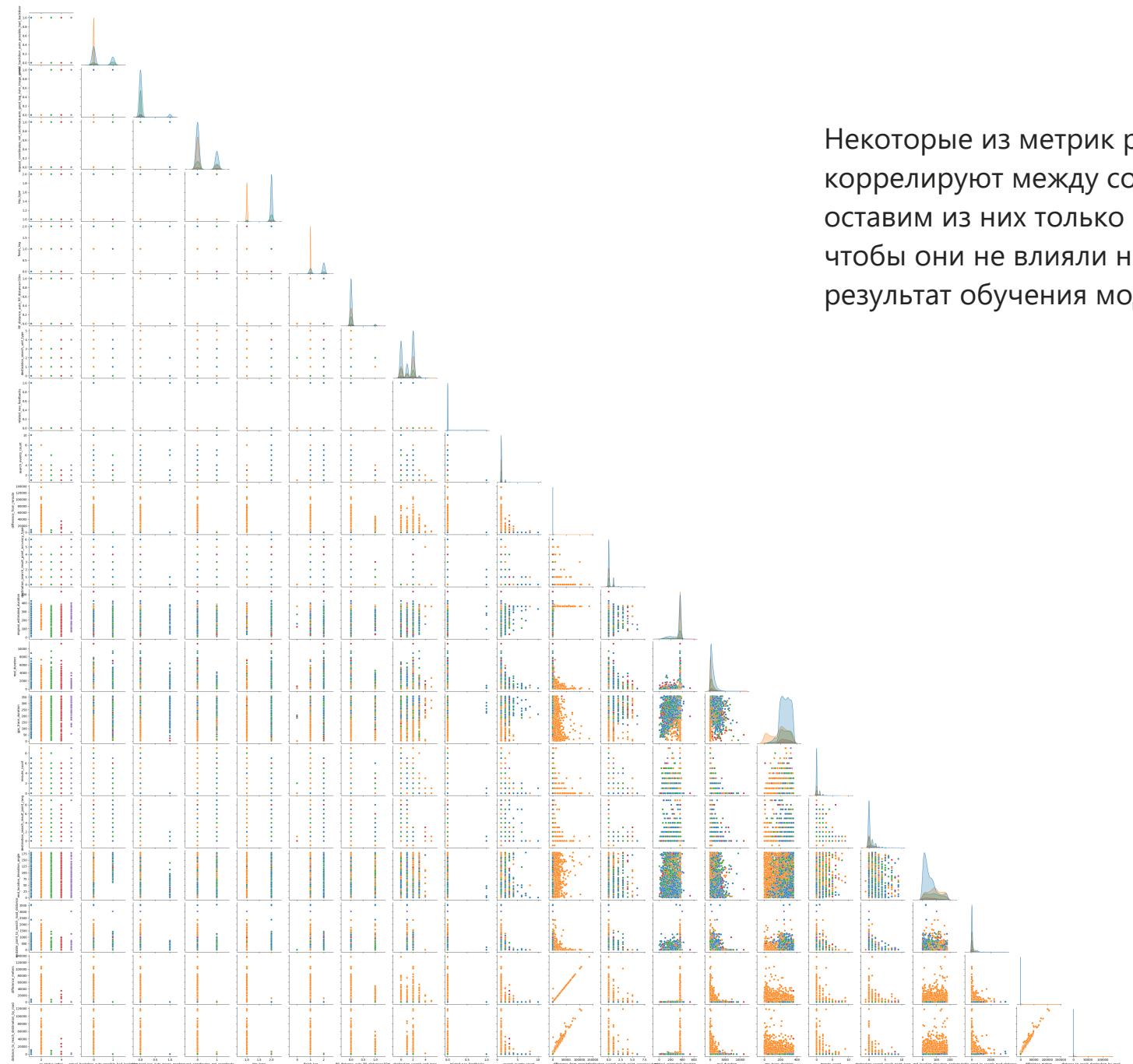
    ▶ Run   Limit 100   Format   View history
145   , current_status, current_tags
146   , case when cardinality(filter(array_distinct(current_tags), x => x like '%3%')) > 0 then 'search_data'
147     when (contains(current_tags,'2:severity_routing_engine_issue_fair') or contains(current_tags,'2:routing_engine_issue')) then 'routing'
148     when (contains(current_tags,'2:severity_search_engine_issue_fair') or contains(current_tags,'2:search_engine_issue')) then 'search_eng'
149     when (contains(current_tags,'2:APP_issue')) or contains(current_tags,'2:severity_APP_issue_fair')) then 'APP_issue'
150     when cardinality(filter(array_distinct(current_tags), x => x like '%1%')) > 0 then 'navigation_data'
151     when cardinality(filter(array_distinct(current_tags), x => x like 'undefined_reason')) > 0 then 'group_undefined_reason'
152     else 'other'
153   end as tag_group
154   , array_distinct(github_issues) github_issues, lat, lon, dt_c, linked_issues
155   from (select *
156   , row_number() over (partition by item_id order by updated_at desc) as rn0
157   , date(created_at) as dt_c
158   from xxxx.xxxxx
159   where project_id = '11111111111111111111111111'
160   and substring(cast(created_at as varchar), 1, 10) between '2025-01-01' and '2025-07-01'
161   where rn0 = 1) m
162 left join yyyy.yyyys f on m.item_id = f.trip_id
163 where rn = []
164 and length(try_cast(reviewed_at as varchar)) > 0
165 order by 1, m.dt_c

```

Данные получаем из базы
данных по следующему SQL
запросу

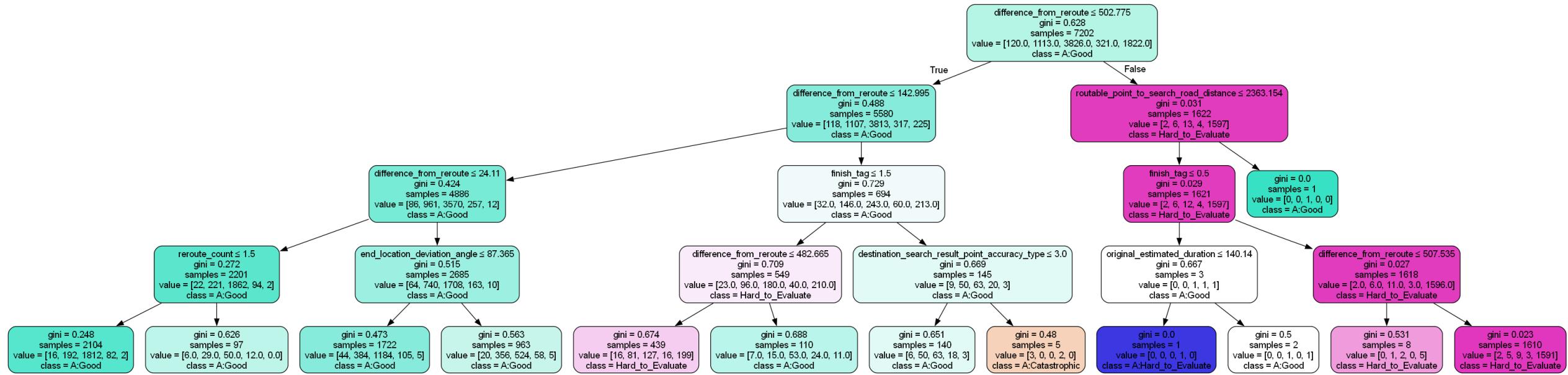
Расстояния при полученном статусе - Hard_to_Evaluate между точкой поиска и точкой прибытия - огромны. На них и сделаем упор в планируемой автоматизации.





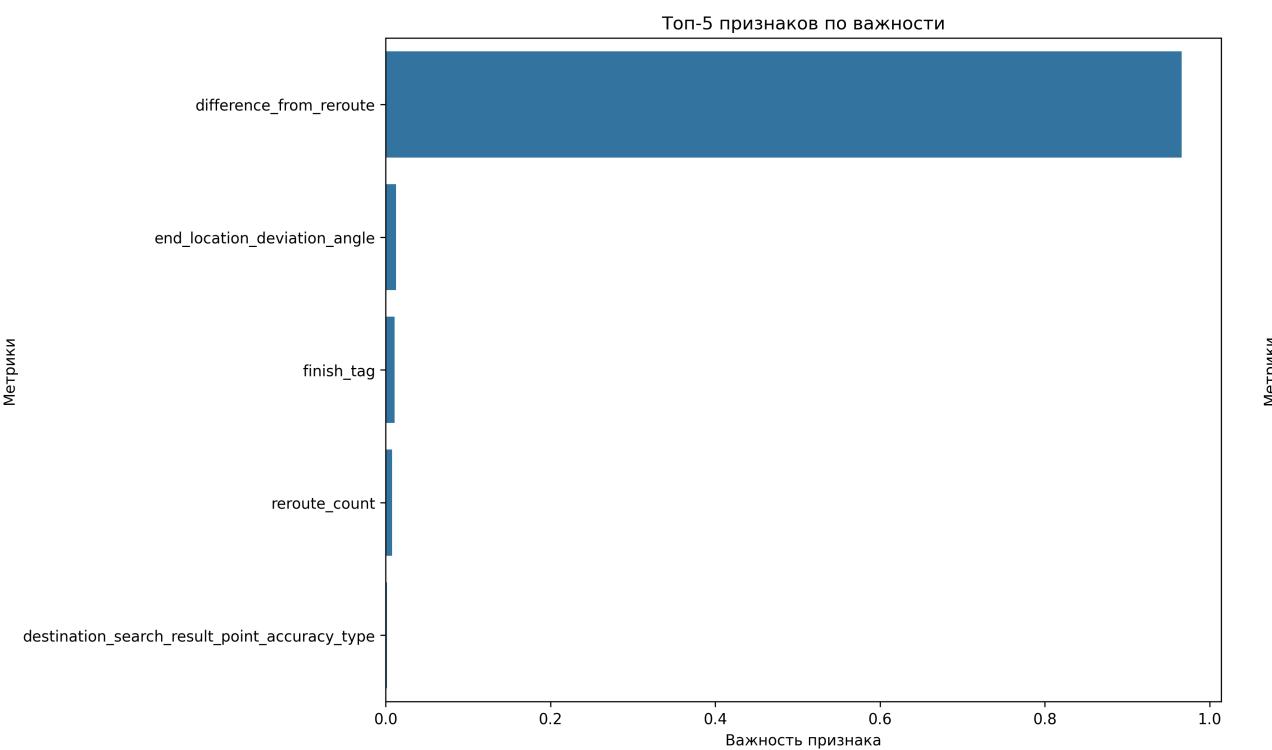
Некоторые из метрик расстояний коррелируют между собой - оставим из них только одну, чтобы они не влияли на результат обучения моделей.

Предположения были верны - одной из решающих конечный статус метрик является метрика расстояния между точкой прибытия и последней точкой перестроения маршрута.

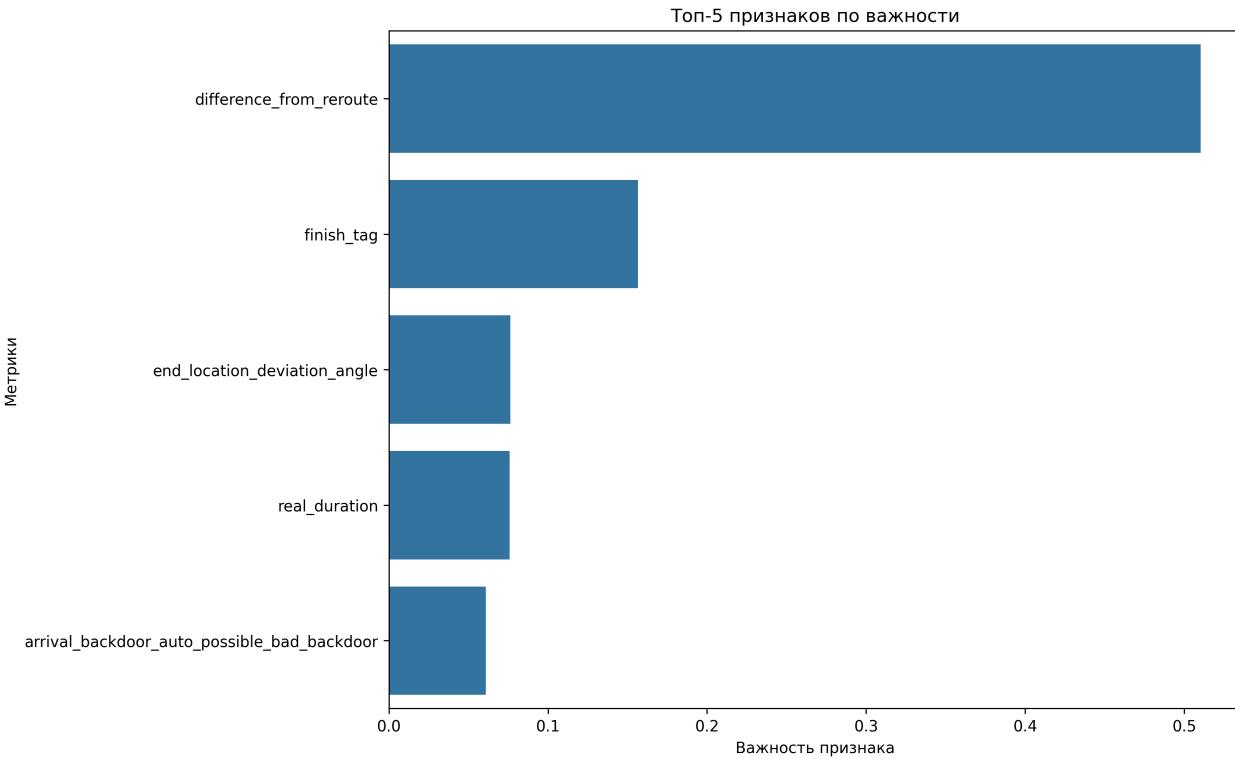


Обе модели с достаточной точностью в этом солидарны. Проверим расчетами.

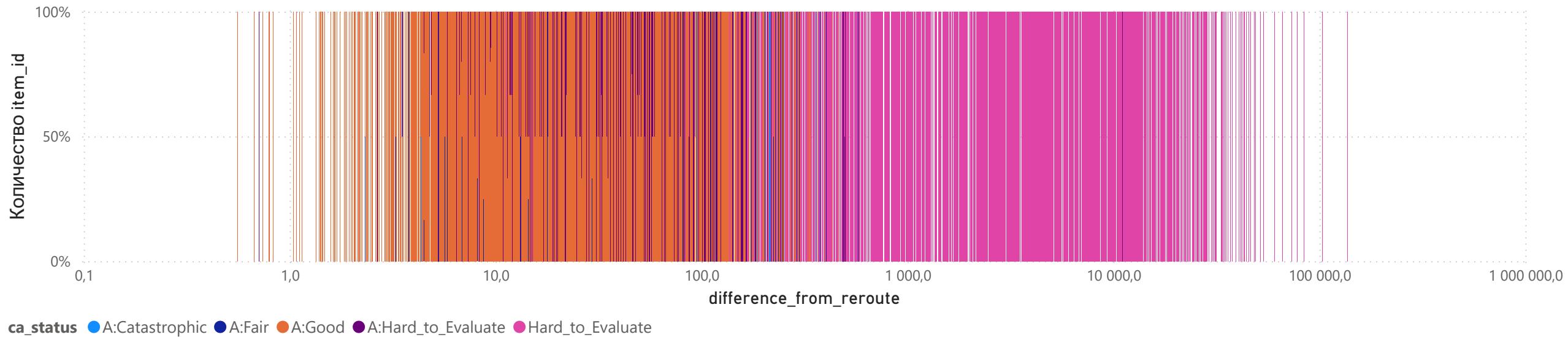
Дерево решений



Лес случайностей



Количество item_id по difference_from_reroute И ca_status



Количество item_id по automated

20% из всех айтемов поездок с
99% точностью можно
автоматизировать, добавляя
статус Hard_to_Evaluate на
предразметке и значительно
уменьшив мануальный
просмотр, исключая их из
датасетов.



