SABINE - SBI Analytics task

This task enables an end-to-end assessment of a whole SBI process starting from clips, possibly enriched with different combinations of the benchmark components.

Challenge

Find the answers to a set of 10 inquiries proposed by our domain experts. The outcome of different enrichment tasks is required to find the correct answers.

I_1: Which is the most discussed sector in relationship with each political party?

I_2: Which are the words most frequently associated to each politician on Twitter?

I_3: Which are the most frequently discussed topics for each source type?

I_4: How frequently are foreign politicians mentioned in each channel supertype?

I_5: How is the discussion of policies differentiated amongst male and female authors?

I_6: Are there any topics whose volume of discussion significantly changes from UK to Italy? [1]

I_7: What is the average sentiment expressed on EPGs and related Parties across different source types?

I_8: How are policies discussed in domestic, European, and foreign perspective?

I_9: How does the sentiment about each politician and technocrat change along time?

I_{10}: What is the percentage of agreement between crowd and expert sentiment for each channel type?

Queries

Query results are shown in the Inquiries.xlsx file.

I1

```
select *
from (
  select dtt1.party party, dts2.name sector, sum(ft.occ) occ,
    rank() over (partition by dtt1.party order by sum(ft.occ) desc) sector_rank_within_party
  from dt_topic dtt1, dt_entity dte1, ft_relcontain ft, dt_entity dte2, bt_policy_sector btps2, dt_sector dts2
  where dtt1.id = dte1.topic_id and dte1.id = ft.first_entity_id and ft.second_entity_id = dte2.id
    and dte2.topic_id = btps2.policy_id and btps2.sector_id = dts2.id
    and dtt1.name <> 'Unclassified'
  group by dtt1.party, dts2.name
) x
where sector_rank_within_party = 1
order by 1;
```

I2

```
select *
from (
  select dtt1.name politician, dte2.name entity, dte2.pos, sum(ft.occ) occ,
    dense_rank() over (partition by dtt1.name order by sum(ft.occ) desc) entity_rank_within_politician
  from dt_topic dtt1, dt_entity dte1, ft_relcontain ft, dt_entity dte2, dt_clip dtc
  where dtt1.politician <> 'Unclassified' and dtt1.id = dte1.topic_id and dte1.id = ft.first_entity_id
```

[1] Note that this is a cross-language enquiry and requires to find out those topics whose number of occurrences are most unbalanced between Italian and English clips

```
    and ft.second_entity_id = dte2.id
    and ft.clip_id = dtc.id and dtc.source = 'twitter.com'
  group by dtt1.name, dte2.name, dte2.pos
) x
where entity_rank_within_politician <= 3
order by 1,2;
```

I3

```
select *
from (
  select dtt.name topic, dtc.source_type, sum(bt.occ) occ,
    rank() over (partition by source_type order by sum(bt.occ) desc) topic_rank_within_source_type
  from dt_topic dtt, bt, dt_clip dtc
  where dtt.id = bt.topic_id and bt.clip_id = dtc.id
  group by dtt.name, dtc.source_type
) x
where topic_rank_within_source <= 5
order by 2,4;
```

I4

```
select channel_supertype, politician_classification, sum(occ) occ
from (
  select dtc.channel_supertype,
    case
      when domestic_yn_of_politician = 'y' then 'domestic'
      when eu_yn_of_politician = 'y' then 'european'
      else 'foreign'
    end politician_classification, bt.occ
  from dt_topic dtt, bt, dt_clip dtc
  where dtt.politician <> 'Unclassified' and dtt.id = bt.topic_id and bt.clip_id = dtc.id
)
group by channel_supertype, politician_classification
order by 1,2;
```

I5

```
select policy, listagg(author_gender,';') within group (order by author_gender) author_genders,
  listagg(occ,';') within group (order by author_gender) occs
from (
  select dtt.policy, dtc.author_gender, sum(bt.occ) occ
  from dt_topic dtt, bt, dt_clip dtc
  where dtt.policy <> 'Unclassified' and dtt.id = bt.topic_id
    and bt.clip_id = dtc.id and author_gender <> 'Unclassified'
  group by dtt.policy, dtc.author_gender
) x
group by policy
order by 1;
```

I6

```
select en.id || '/' || it.id ids, en.name || '/' || it.name names, en.occ occ_en, it.occ occ_it,
  case
    when en.occ/(en.occ+it.occ) > 0.75 then '>> UK'
    when it.occ/(en.occ+it.occ) > 0.75 then '>> Italy'
    else '≈'
  end discussion_volume_comparison
from topic_mapping mp,
  ( select dtt.id, dtt.name, sum(bt.occ) occ
```

```
   from dt_topic dtt, bt
   where dtt.id = bt.topic_id
   group by dtt.id, dtt.name
 ) en,
 ( select dtt.id, dtt.name, sum(bt.occ) occ
   from ita.dt_topic dtt, ita.bt
   where dtt.id = bt.topic_id
   group by dtt.id, dtt.name
 ) it
where it.id = mp.ita_topic_id and en.id = mp.eng_topic_id
 and mp.mapping_semantics = 'SAME-AS'
order by en.occ/(en.occ+it.occ);
```

I7

```
select dtt.epg, dtc.author_country, round(avg(ft.nlp_sentiment),2) avg_sentiment
from dt_topic dtt, bt, ft_raw_clip ft, dt_clip dtc
where dtt.epg <> 'Unclassified' and dtt.politician = 'Unclassified'
 and dtt.id = bt.topic_id and bt.clip_id = ft.clip_id and ft.clip_id = dtc.id
group by dtt.epg, dtc.author_country
order by 1,2;
```

I8

```
select policy, policy_perspective, sum(occ) occ
from (
 select dtt1.policy,
   case
      when dtt2.domestic_yn_of_politician = 'y' or dtt2.domestic_yn_of_party = 'y'
        or dtt2.domestic_yn_of_institution = 'y' or dtt2.domestic_yn_of_technocrat = 'y' then 'domestic'
      when dtt2.eu_yn_of_politician = 'y' or dtt2.eu_yn_of_party = 'y'
        or dtt2.eu_yn_of_institution = 'y' or dtt2.eu_yn_of_technocrat = 'y' then 'european'
      else 'foreign'
    end policy_perspective, ft.occ
 from dt_topic dtt1, dt_entity dte1, ft_relcontain ft, dt_entity dte2, dt_topic dtt2
 where dtt1.id = dte1.topic_id and dte1.id = ft.first_entity_id
   and ft.second_entity_id = dte2.id and dte2.topic_id = dtt2.id
   and (dtt2.domestic_yn_of_politician <> 'Unclassified' or dtt2.domestic_yn_of_party <> 'Unclassified'
        or dtt2.domestic_yn_of_institution <> 'Unclassified' or dtt2.domestic_yn_of_technocrat <> 'Unclassified')
) x
group by policy, policy_perspective
order by 1,2;
```

I9

```
select politician_or_technocrat, name, fulldate, avg(nlp_sentiment) avg_sentiment
from (
 select case when dtt.politician <> 'Unclassified' then 'Politician' else 'Technocrat' end politician_or_technocrat,
   dtt.name, dtd.fulldate, ft.nlp_sentiment
 from dt_date dtd, ft_raw_clip ft, bt, dt_topic dtt
 where dtd.id = ft.date_id and ft.clip_id = bt.clip_id and bt.topic_id = dtt.id
   and (dtt.politician <> 'Unclassified' or dtt.technocrat <> 'Unclassified')
) x
group by politician_or_technocrat, fulldate, name
order by 2,3;
```

I10

```
select channel_type, round(sum(agreement)/count(*),2) agreement_percentage
from (
```

```
  select dtc.channel_type channel_type,
    case
      when ft.expert_sentiment = ft.crowd_sentiment then 1
      else 0
    end agreement
  from dt_clip dtc, ft_raw_clip ft
  where dtc.id = ft.clip_id and ft.expert_sentiment is not null and ft.crowd_sentiment is not null
) x
group by channel_type
order by 1;
```