84factors-ex-sus

October 31, 2017

0.1 WWI, Factor Group and 84 Factors calculation using data excluding suspicious data

```
In [1]: from utility import db_connect, query2csv
        from settings import  DBNAME, DBPASS, DBUSER, DBHOST
```

0.1.1 WWI excluding suspicious data

```
In [2]: qsql="""
        with d as (
          select generate_series(0,6) as dayofweek
        ),
        m as (
          select generate_series(1,12) as month
        ),
        -- v_ijmy:Compute an average by day of week for each month.
        v_ijmy as (
          select
              baadv.analysis_area_id,
              to_char(baadv.date, 'YYYY') as year,
              avg(baadv.volume)::bigint as volume_i,
              avg(baadv.volume) as volume,
              d.dayofweek,
              m.month
          from
              baa_ex_sus.analysis_areas_daily_volume as baadv,
              d,
              m
          where
              extract(dow from baadv.date) in (d.dayofweek)
              AND date_part('month', baadv.date) = m.month
              group by baadv.analysis_area_id, year, d.dayofweek, m.month
        ),
        -- madt: average volume each month, each year for sites
        madt as (
          select
              analysis_area_id,
              month,
```

```
      year,
      avg(volume)::bigint as volume_i,
      avg(volume) as volume
  from
      v_ijmy
      group by analysis_area_id, year, month
      having count(dayofweek)=7 -- having 7 days of data each week
),
AADT as (
select
  analysis_area_id,
  year,
  avg(volume)::bigint as AADT_i,
  round(avg(volume), 2) as AADT
from madt
  group by analysis_area_id, year
  having count(month) = 12 -- having 12 months of data
),
-- daily_exclude_holiday: daily counts for sites excluding holidays
daily_exclude_holiday as (
select
 baaad.analysis_area_id,
 baaad.date,
 baaad.volume,
 date_part('month', baaad.date) as month,
 date_part('dow', baaad.date) as dow
from
  baa_ex_sus.analysis_areas_daily_volume as baaad
  left join baa.holidays as baahd on baaad.date::date = baahd.holiday_date
where
  baahd.holiday_id is null
  group by 1,2,3
),
V_jmyl_exclude_holiday as (
  select
      baadv.analysis_area_id,
      to_char(baadv.date, 'YYYY') as year,
      avg(baadv.volume) as volume,
      d.dayofweek,
      m.month
  from
      daily_exclude_holiday as baadv,
      d,
      m
  where
      extract(dow from baadv.date) in (d.dayofweek)
      AND date_part('month', baadv.date) = m.month
      group by baadv.analysis_area_id, year, d.dayofweek, m.month
```

```
),
-- 84 factors volume count should exclude holiday weeks
factor84 as (
select
  v_jmyl_nh.analysis_area_id,
  v_jmyl_nh.volume as v_jmyl,
  AADT.aadt as aadt,
  round(v_jmyl_nh.volume/aadt::numeric, 2) as f_jmys,
  v_jmyl_nh.dayofweek,
  v_jmyl_nh.month,
  v_jmyl_nh.year
from
  V_jmyl_exclude_holiday as v_jmyl_nh inner join AADT using(analysis_area_i
where
  AADT.AADT <> 0
),
V_we as (
select
  baadv.analysis_area_id,
  avg(baadv.volume) vwe
from
  baa_ex_sus.analysis_areas_daily_volume as baadv
where
  extract(dow from baadv.date) in (0,6)
  group by baadv.analysis_area_id
),
V_wd as (
select
  baadv.analysis_area_id,
  avg(baadv.volume) vwd
from
  baa_ex_sus.analysis_areas_daily_volume as baadv
where
  extract(dow from baadv.date) in (1,2,3,4,5)
  group by baadv.analysis_area_id
),
grouping as (
select
  V_we.analysis_area_id,
  round(V_we.vwe, 2) as V_we,
  round(V_wd.vwd, 2) as V_wd,
  round(V_we.vwe/V_wd.vwd, 2) as wwi,
  case
     when (round(V_we.vwe/V_wd.vwd, 2) <= 0.8) then 'Weekday Commute'
     when (round(V_we.vwe/V_wd.vwd, 2) >  1.2) then 'Weekend Multipurpose'
     ELSE 'Weekly Multipurpose'
  END as grouping
from
```

```
  V_we inner join V_wd using (analysis_area_id)
),
wwi as (
select
  grouping.analysis_area_id,
  baaa.mode,
  baaa.analysis_area_name,
  grouping.v_we,
  grouping.v_wd,
  grouping.wwi,
  grouping.grouping as weekly_group
from
   grouping inner join baa.analysis_areas as baaa using(analysis_area_id)
   order by 1
)
select * from wwi
"""
csvfile='wwi-ex-sus.csv'
query2csv(qsql,csvfile)
```

```
<IPython.core.display.HTML object>
```

0.1.2 Factor group excluding suspicious data

```
In [3]: qsql="""
with V_we as (
select
  baadv.analysis_area_id,
  avg(baadv.volume) vwe
from
  baa_ex_sus.analysis_areas_daily_volume as baadv
where
  extract(dow from baadv.date) in (0,6)
  group by baadv.analysis_area_id
),
V_wd as (
select
  baadv.analysis_area_id,
  avg(baadv.volume) vwd
from
  baa_ex_sus.analysis_areas_daily_volume as baadv
where
  extract(dow from baadv.date) in (1,2,3,4,5)
  group by baadv.analysis_area_id
),
pre_wwi as (
select
```

```
  V_we.analysis_area_id,
  round(V_we.vwe, 2) as V_we,
  round(V_wd.vwd, 2) as V_wd,
  round(V_we.vwe/V_wd.vwd, 2) as wwi,
  case
     when (round(V_we.vwe/V_wd.vwd, 2) <= 0.8) then 'Weekday Commute'
     when (round(V_we.vwe/V_wd.vwd, 2) >  1.2) then 'Weekend Multipurpose'
     ELSE 'Weekly Multipurpose'
  END as grouping
from
  V_we inner join V_wd using (analysis_area_id)
),
wwi as (
select
  wwi.analysis_area_id,
  baaa.mode,
  baaa.analysis_area_name,
  baaa.analysis_area_regions_id,
  wwi.v_we,
  wwi.v_wd,
  wwi.wwi,
  wwi.grouping as weekly_group
from
   pre_wwi as wwi inner join baa.analysis_areas as baaa using(analysis_area
   order by 1
),
factorgrp as (
  select
    ar.analysis_area_name as city,
    wwi.mode,
    wwi.weekly_group,
    array_agg(wwi.analysis_area_id order by analysis_area_id) as analysis_a
  from
    wwi, baa.analysis_area_regions as ar
  where
    ar.analysis_area_regions_id = wwi.analysis_area_regions_id
    group by 1,2,3
)
select * from   factorgrp
order by 1,2,3
"""
csvfile='factor-group-wwi-ex-sus.csv'
query2csv(qsql,csvfile)
```

```
<IPython.core.display.HTML object>
```

0.1.3 Creating factor group table

```
CREATE TABLE baa_ex_sus.factor_group
(
  city character varying(50) NOT NULL,
  weekly_group character varying(50) NOT NULL,
  mode baa.bp_mode NOT NULL,
  analysis_area_id_list integer[]
)
```

0.1.4 Populate factor group table

```
insert into baa_ex_sus.factor_group
(
with V_we as (
select
  baadv.analysis_area_id,
  avg(baadv.volume) vwe
from
  baa_ex_sus.analysis_areas_daily_volume as baadv
where
  extract(dow from baadv.date) in (0,6)
  group by baadv.analysis_area_id
),
V_wd as (
select
  baadv.analysis_area_id,
  avg(baadv.volume) vwd
from
  baa_ex_sus.analysis_areas_daily_volume as baadv
where
  extract(dow from baadv.date) in (1,2,3,4,5)
  group by baadv.analysis_area_id
),
pre_wwi as (
select
  V_we.analysis_area_id,
  round(V_we.vwe, 2) as V_we,
  round(V_wd.vwd, 2) as V_wd,
  round(V_we.vwe/V_wd.vwd, 2) as wwi,
  case
     when (round(V_we.vwe/V_wd.vwd, 2) <= 0.8) then 'Weekday Commute'
     when (round(V_we.vwe/V_wd.vwd, 2) >  1.2) then 'Weekend Multipurpose'
     ELSE 'Weekly Multipurpose'
  END as grouping
from
  V_we inner join V_wd using (analysis_area_id)
),
```

```
wwi as (
select
  wwi.analysis_area_id,
  baaa.mode,
  baaa.analysis_area_name,
  baaa.analysis_area_regions_id,
  wwi.v_we,
  wwi.v_wd,
  wwi.wwi,
  wwi.grouping as weekly_group
from
   pre_wwi as wwi inner join baa.analysis_areas as baaa using(analysis_area_id)
   order by 1
),
factorgrp as (
  select
    ar.analysis_area_name as city,
    wwi.mode,
    wwi.weekly_group,
    array_agg(wwi.analysis_area_id order by analysis_area_id) as analysis_area_id_l
  from
    wwi, baa.analysis_area_regions as ar
  where
    ar.analysis_area_regions_id = wwi.analysis_area_regions_id
    group by 1,2,3
)
select city, weekly_group, mode, analysis_area_id_list from   factorgrp
order by 1,2,3
)
```

```
In [5]: qsql="""
        select * from baa_ex_sus.factor_group
        """
        csvfile='baa_ex_sus_factor_group.csv'
        query2csv(qsql,csvfile)
```

```
<IPython.core.display.HTML object>
```

0.1.5 84 Factors excluding suspicious data

```
In [6]: qsql="""
        with d as (
          select generate_series(0,6) as dayofweek
        ),
        m as (
          select generate_series(1,12) as month
        ),
```

```
-- v_ijmy:Compute an average by day of week for each month.
v_ijmy as (
  select
      baadv.analysis_area_id,
      to_char(baadv.date, 'YYYY') as year,
      avg(baadv.volume)::bigint as volume_i,
      avg(baadv.volume) as volume,
      d.dayofweek,
      m.month
  from
      baa_ex_sus.analysis_areas_daily_volume as baadv,
      d,
      m
  where
      extract(dow from baadv.date) in (d.dayofweek)
      AND date_part('month', baadv.date) = m.month
      group by baadv.analysis_area_id, year, d.dayofweek, m.month
),
-- madt: average volume each month, each year for sites
madt as (
  select
      analysis_area_id,
      month,
      year,
      avg(volume)::bigint as volume_i,
      avg(volume) as volume
  from
      v_ijmy
      group by analysis_area_id, year, month
      having count(dayofweek)=7 -- having 7 days of data each week
),
AADT as (
select
  analysis_area_id,
  year,
  avg(volume)::bigint as AADT_i,
  round(avg(volume), 2) as AADT
from madt
  group by analysis_area_id, year
  having count(month) = 12 -- having 12 months of data
),
-- daily_exclude_holiday: daily counts for sites excluding holidays
daily_exclude_holiday as (
select
 baaad.analysis_area_id,
 baaad.date,
 baaad.volume,
 date_part('month', baaad.date) as month,
```

```
 date_part('dow', baaad.date) as dow
from
  baa_ex_sus.analysis_areas_daily_volume as baaad
  left join baa.holidays as baahd on baaad.date::date = baahd.holiday_date
where
  baahd.holiday_id is null
  group by 1,2,3
),
V_jmyl_exclude_holiday as (
  select
      baadv.analysis_area_id,
      to_char(baadv.date, 'YYYY') as year,
      avg(baadv.volume) as volume,
      d.dayofweek,
      m.month
  from
      daily_exclude_holiday as baadv,
      d,
      m
  where
      extract(dow from baadv.date) in (d.dayofweek)
      AND date_part('month', baadv.date) = m.month
      group by baadv.analysis_area_id, year, d.dayofweek, m.month
),
-- 84 factors volume count should exclude holiday weeks
factor84 as (
select
  v_jmyl_nh.analysis_area_id,
  v_jmyl_nh.volume as v_jmyl,
  AADT.aadt as aadt,
  round(v_jmyl_nh.volume/aadt::numeric, 2) as f_jmys,
  v_jmyl_nh.dayofweek,
  v_jmyl_nh.month,
  v_jmyl_nh.year
from
  V_jmyl_exclude_holiday as v_jmyl_nh inner join AADT using(analysis_area_i
where
  AADT.AADT <> 0
),
V_we as (
select
  baadv.analysis_area_id,
  avg(baadv.volume) vwe
from
  baa_ex_sus.analysis_areas_daily_volume as baadv
where
  extract(dow from baadv.date) in (0,6)
  group by baadv.analysis_area_id
```

```
),
V_wd as (
select
  baadv.analysis_area_id,
  avg(baadv.volume) vwd
from
  baa_ex_sus.analysis_areas_daily_volume as baadv
where
  extract(dow from baadv.date) in (1,2,3,4,5)
  group by baadv.analysis_area_id
),
grouping as (
select
  V_we.analysis_area_id,
  round(V_we.vwe, 2) as V_we,
  round(V_wd.vwd, 2) as V_wd,
  round(V_we.vwe/V_wd.vwd, 2) as wwi,
  case
     when (round(V_we.vwe/V_wd.vwd, 2) <= 0.8) then 'Weekday Commute'
     when (round(V_we.vwe/V_wd.vwd, 2) >  1.2) then 'Weekend Multipurpose'
     ELSE 'Weekly Multipurpose'
  END as grouping
from
  V_we inner join V_wd using (analysis_area_id)
),
wwi as (
select
  grouping.analysis_area_id,
  baaa.mode,
  baaa.analysis_area_name,
  baaa.analysis_area_regions_id,
  grouping.v_we,
  grouping.v_wd,
  grouping.wwi,
  grouping.grouping as weekly_group
from
   grouping inner join baa.analysis_areas as baaa using(analysis_area_id)
),
-- group analysis_area_id into array
factorgrp as (
 select
    ar.analysis_area_name as city,
    wwi.mode,
    wwi.weekly_group,
    array_agg(wwi.analysis_area_id order by analysis_area_id) as analysis_a
  from
    wwi, baa.analysis_area_regions as ar
  where
```

```
            ar.analysis_area_regions_id = wwi.analysis_area_regions_id
            group by 1,2,3
          )
          select
            fg.city,
            fg.weekly_group,
            fg.mode,
            fg.analysis_area_id_list,
            f84.dayofweek,
            f84.month,
            f84.year,
            round(avg(f84.f_jmys), 2) as f_jmys_avg
          from
            factor84 as f84 inner join factorgrp as fg
            on f84.analysis_area_id = Any(fg.analysis_area_id_list::int[])
            group by
            fg.city,
            fg.weekly_group,
            fg.mode,
            fg.analysis_area_id_list,
            f84.dayofweek,
            f84.month,
            f84.year
            order by fg.city,
            fg.weekly_group,
            fg.mode,f84.year, f84.month, f84.dayofweek
        """
        csvfile='84factor_groups-ex-sus.csv'
        query2csv(qsql,csvfile)
```

```
<IPython.core.display.HTML object>
```