

Eksplorasi Data Harga Pangan Indonesia

Dony Indiarto

2022-12-06

Data harga pangan

Data harga pangan merupakan salah satu masukan untuk menghitung ongkos pangan sehat di suatu waktu dan tempat yang spesifik.

Data harga pangan diperoleh dari data portal HDX dengan data versi tanggal 27 November 2022. Data tersebut merupakan irisan dari database Harga Pangan World Food Programme (WFP) yang mencakup jenis makanan seperti jagung, beras, kacang-kacangan, ikan, dan gula untuk 98 negara dan sekitar 3000 pasar.

Eksplorasi data

Dimensi data

```
fp <- read_csv("data/wfp_food_prices_idn.csv")
```

```
## Rows: 112822 Columns: 14
## -- Column specification -----
## Delimiter: ","
## chr (14): date, admin1, admin2, market, latitude, longitude, category, commo...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
glimpse(fp)
```

```
## Rows: 112,822
## Columns: 14
## $ date      <chr> "date", "2007-01-15", "2007-01-15", "2007-01-15", "2007-01-15"
## $ admin1    <chr> "adm1+name", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA
## $ admin2    <chr> "adm2+name", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA
## $ market    <chr> "loc+market+name", "National Average", "National Average", "National Average"
## $ latitude  <chr> "geo+lat", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA
## $ longitude <chr> "geo+lon", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA
## $ category  <chr> "item+type", "cereals and tubers", "cereals and tubers", "cereals and tubers"
## $ commodity <chr> "item+name", "Rice", "Wheat flour", "Eggs", "Meat (beef)", "Meat (beef)", "Meat (beef)"
## $ unit      <chr> "item+unit", "KG", "KG", "KG", "KG", "KG", "385 G", "KG", "KG", "KG", "KG", "KG", "KG"
```

```
## $ priceflag <chr> "#item+price+flag", "actual", "actual", "actual", "actual", ~
## $ pricetype <chr> "#item+price+type", "Retail", "Retail", "Retail", "Retail", ~
## $ currency <chr> "#currency", "IDR", "IDR", "IDR", "IDR", "IDR", "IDR", "IDR"~
## $ price <chr> "#value", "5941.98", "4328.26", "9150.74", "49165.11", "2012~
## $ usdprice <chr> "#value+usd", "0.653", "0.4756", "1.0056", "5.4029", "2.2114~
```

Cakupan temporal

```
# Jumlah bulan ketersediaan data
fp |>
  select(date) |>
  distinct() |>
  nrow()
```

```
## [1] 172
```

```
# Sebaran ketersediaan data
fp |>
  select(date) |>
  distinct() |>
  slice(-1) |>
  mutate(tag = 1) |>
  mutate(date = as_date(date),
          month = month(date),
          year = year(date)) |>
  pivot_wider(id_cols = month, values_from = tag, names_from = year)
```

```
## # A tibble: 12 x 16
##   month '2007' '2008' '2009' '2010' '2011' '2012' '2013' '2014' '2015' '2016'
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1     1     1     1     1     1     1     1     1     1     1
## 2     2     1     1     1     1     1     1     1     1     1     1
## 3     3     1     1     1     1     1     1     1     1     1     1
## 4     4     1     1     1     1     1     1     1     1     1     1
## 5     5     1     1     1     1     1     1     1     1     1     1
## 6     6     1     1     1     1     1     1     1     1     1     1
## 7     7     1     1     1     1     1     1     1     1     1     1
## 8     8     1     1     1     1     1     1     1     1     1     1
## 9     9     1     1     1     1     1     1     1     1     1     1
## 10    10     1     1     1     1     1     1     1     1     1     1
## 11    11     1     1     1     1     1     1     1     1     1     1
## 12    12     1     1     1     1     1     1     1     1     1     1
## # ... with 5 more variables: '2017' <dbl>, '2018' <dbl>, '2019' <dbl>,
## #   '2020' <dbl>, '2021' <dbl>
```

Aspek keterwakilan spasial

Provinsi dan Kabupaten/Kota yang tercakup beserta jumlah pasarnya

```
cakupan_wilayah <- fp |>
  slice(-1) |>
  filter(market!="National Average") |>
  select(admin1, admin2, market) |> distinct() |>
  group_by(admin1, admin2) |> tally()

print(cakupan_wilayah)
```

```
## # A tibble: 106 x 3
## # Groups:   admin1 [34]
##   admin1          admin2      n
##   <chr>          <chr>    <int>
## 1 ACEH          ACEH BARAT      2
## 2 ACEH          KOTA BANDA ACEH  2
## 3 ACEH          KOTA LHOKESEUMAWA 2
## 4 BALI          BULELENG      2
## 5 BALI          KOTA DENPASAR   2
## 6 BANTEN        KOTA CILEGON    2
## 7 BANTEN        KOTA SERANG     2
## 8 BANTEN        KOTA TANGERANG  2
## 9 BENGKULU      KOTA BENGKULU   4
## 10 DAERAH ISTIMEWA YOGYAKARTA KOTA YOGYAKARTA  2
## # ... with 96 more rows
```

Data mencakup 34 Provinsi, 106 Kabupaten/Kota, dan 212 pasar daerah.

```
cakupan_wilayah |> pull(n) |> sum()
```

```
## [1] 212
```

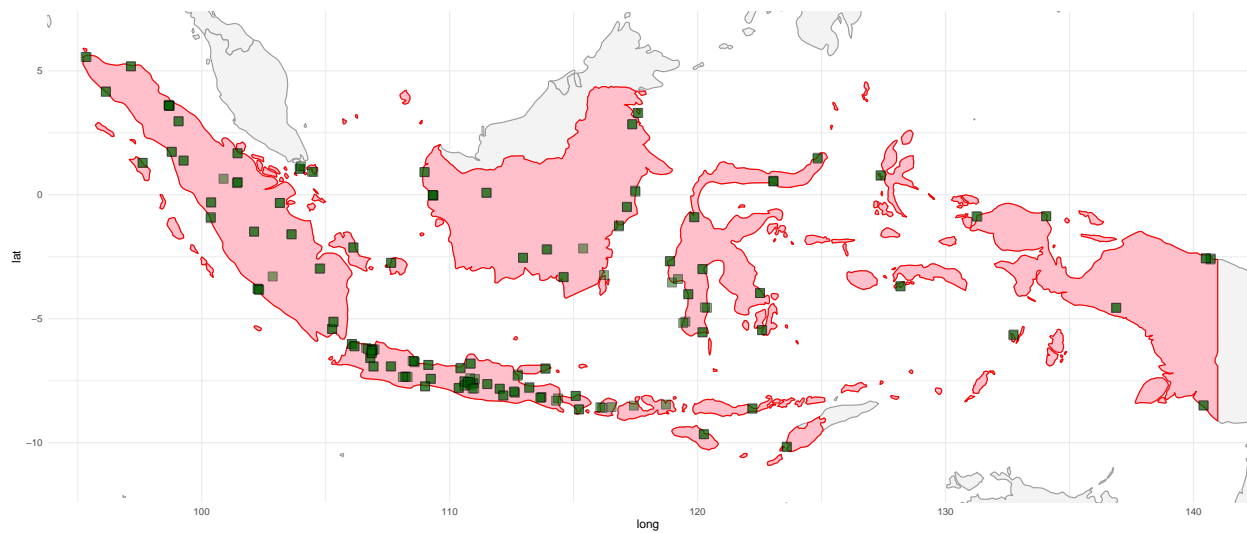
```
map_ina <- map_data('world')[map_data('world')$region == "Indonesia",]

market_ina <-fp |>
  slice(-1) |>
  filter(market!="National Average") |>
  select(admin1, admin2, market, latitude, longitude) |>
  distinct() |> mutate(longitude = as.numeric(longitude),
                      latitude = as.numeric(latitude)) #/>
  #sf::st_as_sf( coords = c("longitude", "latitude"),
  #             crs = 4326, agr = "constant")

## The map (maps + ggplot2 )
ggplot() +
  ## First layer: worldwide map
  geom_polygon(data = map_data("world"),
              aes(x=long, y=lat, group = group),
              color = '#9c9c9c', fill = '#f3f3f3') +
  ## Second layer: Country map
  geom_polygon(data = map_ina,
              aes(x=long, y=lat, group = group),
              color = 'red', fill = 'pink') +
```

```
geom_point(data = market_ina,
           aes(x=longitude, y=latitude, group = market), size = 4, shape = 22, fill = "darkgreen", al
coord_map()+
coord_fixed(
           xlim = c( 96, 140),
           ylim = c(-11.5, 6.5)) +
theme_minimal()
```

```
## Coordinate system already present. Adding new coordinate system, which will
## replace the existing one.
```



Data kosong