**Visualisasi Data Statis dengan Dataset Kesejahteraan Pekerja Indonesia**



**Disusun sebagai proyek dari mata kuliah**

**IF4061 Visualisasi Data dan Informasi**

**oleh**

**Fayza Nadia 13520001**

**Ahmad Alfani Handoyo 13520023**

**Owen Christian Wijaya 13520124**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**SEKOLAH TEKNIK ELEKTRO & INFORMATIKA**

**INSTITUT TEKNOLOGI BANDUNG**

**MARET 2024**

# DAFTAR ISI

[**DAFTAR ISI 1**](#_q10qd4o5ag4o)

[**DAFTAR TABEL 2**](#_jd8gwjvv7js7)

[**BAB I DEFINE PURPOSE AND PARAMETER 3**](#_ir3nek78tinl)

[I.1 Topic 3](#_hseg7wvcmr80)

[I.2 Purpose 3](#_y0lok1lp4ss2)

[I.3 Users 3](#_1o7bds2nnd8b)

[I.4 Intention 4](#_u74uyrolbqdf)

[I.4.1 Function 4](#_txrzq4idnv04)

[I.4.2 Tone 4](#_jdz76rtukrwm)

[**BAB II DATA PREPARATION AND EXPLORATION 5**](#_3nfjdeefr2rq)

[II.1 Acquisition 5](#_a8p8382370i6)

[II.2 Examination & Data Type 5](#_vujtt2gke31r)

[II.2.1 Dataset Upah Pekerja per Jam 5](#_uvdvweo0s4lr)

[II.2.2 Dataset Upah Minimum Provinsi (UMP) 6](#_fwcrus34pyuk)

[II.2.3 Dataset Garis Kemiskinan 6](#_u8l0vs5gt8cr)

[II.2.4 Dataset Pengeluaran per Kapita (PPK) 7](#_arytvaulf07p)

[II.3 Transforming for Quality 8](#_n36w7chjuja5)

[II.4 Transforming for Analysis 8](#_iar332guekfr)

[II.5 Consolidating 8](#_hiyt8hgjtdkz)

[**BAB III QUESTION FORMULATION 9**](#_1fpv4fjqtmdz)

[III.1 Editorial Focus 9](#_wdqy10u3wn0j)

[III.2 Reasoning 9](#_jlqpiw6dbtuy)

[III.3 Narrative 9](#_dmvv7r44eggu)

[III.3.1 Genre 9](#_j7phf78pe76s)

[III.3.2 Visual Narrative Tactics 9](#_m7ynm1a90ju8)

[III.3.3 Narrative Structure Tactics 10](#_kuqjy7pz1yyq)

[**REFERENSI 11**](#_z2gld7j366jb)

# 

# DAFTAR TABEL

# 

[Tabel II.1 Tipe Data dan Deskripsi Tambahan dalam Dataset Upah Pekerja per Jam 5](#_xj3fff9vll5v)

[Tabel II.2 Tipe Data dan Deskripsi Tambahan dalam Dataset UMP 6](#_hqnvrxr1nh49)

[Tabel II.3 Tipe Data dan Deskripsi Tambahan dalam Dataset GK 7](#_6iyl0a3puo3c)

[Tabel II.4 Tipe Data dan Deskripsi Tambahan dalam Dataset PPK 7](#_pnbj88vjxk5i)

# 

# DEFINE PURPOSE AND PARAMETER

## Topic

Bekerja menjadi kegiatan yang esensial bagi seluruh penduduk untuk memenuhi kebutuhan hidup sehari-hari. Namun, pemenuhan kesejahteraan para pekerja menjadi suatu tantangan tersendiri di Indonesia. Setiap provinsi di Indonesia memiliki angka Upah Minimum Provinsi (UMP) dan rata-rata upah pekerja yang berbeda antara satu dengan yang lainnya. Sama halnya dengan upah atau pemasukan, angka pengeluaran serta garis kemiskinan juga beragam berdasarkan daerahnya masing-masing.

Bagi masyarakat yang hendak melakukan urbanisasi ke provinsi tertentu untuk mencari pekerjaan, disarankan untuk mengetahui informasi mengenai pendapatan yang akan diperoleh serta kemampuan pendapatan untuk memenuhi kebutuhan sehari-hari. Informasi tersebut umumnya dapat diperoleh melalui Badan Pusat Statistik (BPS), namun masyarakat umum belum tentu memiliki kemampuan yang diperlukan untuk mengolah data tersebut, memvisualisasikan data, serta mendapatkan wawasan yang diperlukan. Untuk mendukung kebutuhan tersebut, kami mengangkat topik mengenai tren pendapatan dan pengeluaran di provinsi-provinsi Indonesia.

## Purpose

Tujuan yang ingin dicapai adalah:

1. Menyajikan informasi yang mendalam dan komprehensif mengenai kondisi kesejahteraan pekerja di seluruh Indonesia. Dengan demikian, diharapkan masyarakat dapat memahami perbedaan dan tantangan yang dihadapi oleh para pekerja di berbagai provinsi.
2. Memberikan panduan bagi individu yang berencana melakukan urbanisasi ke provinsi tertentu dengan memberikan data terkait pendapatan dan pengeluaran di masing-masing provinsi. Hal ini diharapkan dapat menjadi landasan bagi mereka dalam membuat keputusan strategis terkait pemilihan lokasi kerja.

## Users

Sasaran pengguna dari visualisasi data yang akan dibuat adalah masyarakat umum. Harapannya, visualisasi data dapat diterima dengan mudah oleh pengguna dari berbagai kalangan, di mana pengguna dapat memahami dan mendapatkan informasi dari visualisasi tanpa diperlukannya pengetahuan spesifik pada domain tertentu. Demografi pengguna yang ditujukan dalam visualisasi data ini adalah para pekerja dan calon pekerja di seluruh Indonesia, juga para warga yang hendak melakukan urbanisasi. Melalui tujuan kami, diharapkan para pengguna potensial nantinya akan memperoleh informasi mengenai provinsi yang strategis untuk memenuhi kebutuhan finansial masing-masing berdasarkan sejarah penghasilan dan pengeluaran dari sebuah provinsi.

## Intention

### Function

Fungsionalitas dari visualisasi data adalah *explanatory*, yaitu visualisasi data yang dibuat bersifat spesifik dan terfokus. Sesuai dengan tujuan yang ingin dicapai, visualisasi data diharapkan dapat membuka gambaran bagi pengguna terkait kondisi kesejahteraan pekerja berdasarkan penjelasan dan narasi yang dibangun.

### Tone

Tonalitas dari visualisasi data adalah pragmatis dan analitis. Visualisasi data yang dirancang diharapkan dapat menuai reaksi pengguna yang membaca untuk mengamati dan mengobservasi lebih dalam data yang ditampilkan tersebut secara objektif.

# DATA PREPARATION AND EXPLORATION

## Acquisition

Akuisisi data dilakukan melalui *platform Kaggle*, salah satu *platform* pembelajaran mesin dan *repository* data terbuka yang dapat diakses secara terbuka. Data yang diperoleh adalah hasil pemrosesan data yang bersumber dari Badan Pusat Statistik. Dokumen diperoleh dalam format *comma-separated values* (CSV) yang dapat direpresentasikan dalam bentuk tabular dan dapat diolah lebih lanjut menggunakan pustaka-pustaka dalam bahasa pemrograman Python. Untuk setiap dokumen, terdapat dua buah versi *file* CSV yang diperoleh:

1. File berimbuhan .csv yang merupakan data mentah yang belum melalui proses disagregasi
2. File berimbuhan .df.csv yang merupakan hasil proses disagregasi terhadap data mentah berdasarkan kategori-kategori tertentu

Berikut adalah dokumen-dokumen yang digunakan dan diperoleh dari *platform* tersebut.

1. Data rata-rata upah pekerja per jam dengan disagregasi berdasarkan provinsi dan tahun.
2. Data upah minimum provinsi (UMP) dengan disagregasi berdasarkan provinsi dan tahun.
3. Data garis kemiskinan (GK) per kapita dengan disagregasi berdasarkan provinsi, tahun, periode survei, jenis pengeluaran, dan daerah tempat tinggal.
4. Data rata-rata pengeluaran per kapita (PPK) dengan disagregasi berdasarkan provinsi, tahun, jenis pengeluaran, serta daerah tempat tinggal.

## Examination & Data Type

Pemeriksaan dilakukan menggunakan bahasa Python dan kakas pengolahan data seperti Pandas, NumPy, dan Matplotlib untuk melakukan visualisasi data awal dan pengolahan data. Pemeriksaan dan pemrosesan lebih lanjut akan dilakukan pada dataset hasil disagregasi.

### Dataset Upah Pekerja per Jam

Dataset upah pekerja per jam merepresentasikan data rata-rata upah pekerja minimum dalam satuan Rupiah untuk setiap provinsi di Indonesia pada rentang tahun 2015 hingga 2022. Dari hasil evaluasi, tidak terdapat kekosongan data pada dataset ini, dan terdapat distribusi yang rata pada data untuk setiap provinsi. Pada dataset ini tidak terdapat data duplikat. Berikut adalah tabel mengenai tipe data serta deskripsi tambahan mengenai kolom dalam dataset upah pekerja per jam.

#### Tabel II.1 Tipe Data dan Deskripsi Tambahan dalam Dataset Upah Pekerja per Jam

| **Nama Kolom** | **Tipe Data** | **Deskripsi** |
| --- | --- | --- |
| Provinsi | Kategorikal nominal | 34 provinsi unik Indonesia sebelum tahun 2022. |
| Tahun | Kuantitatif interval | Tahun di saat data diperoleh dalam rentang 2015 hingga 2022. |
| Upah | Kuantitatif rasio | Upah rerata pada suatu daerah dalam satuan Rupiah. |

### Dataset Upah Minimum Provinsi (UMP)

Dataset UMP merepresentasikan data rata-rata upah minimum provinsi dalam satuan Rupiah untuk setiap provinsi di Indonesia pada rentang tahun 2002 hingga 2022. Dari aspek kelengkapan data, terdapat kekosongan nilai UMP pada beberapa provinsi dalam rentang tahun 2002 hingga 2004, kecuali pada provinsi Kalimantan Utara dengan kekosongan nilai UMP dalam rentang tahun 2002 hingga 2014. Pada dataset ini tidak terdapat data duplikat. Berikut adalah tabel mengenai tipe data serta deskripsi tambahan mengenai kolom dalam dataset UMP.

#### Tabel II.2 Tipe Data dan Deskripsi Tambahan dalam Dataset UMP

| **Nama Kolom** | **Tipe Data** | **Deskripsi** |
| --- | --- | --- |
| Provinsi | Kategorikal nominal | 34 provinsi unik Indonesia sebelum tahun 2022. |
| Tahun | Kuantitatif interval | Tahun di saat data diperoleh dalam rentang 2002 hingga 2022. |
| UMP | Kuantitatif rasio | Upah minimum provinsi pada suatu daerah dalam satuan Rupiah. |

### Dataset Garis Kemiskinan

Dataset ini merepresentasikan garis kemiskinan untuk setiap provinsi dalam rentang tahun 2014 hingga 2022. Setiap data garis kemiskinan diambil pada bulan Maret dan September pada setiap tahun di antara rentang tersebut. Terdapat tiga buah jenis garis kemiskinan, yaitu garis kemiskinan berdasarkan kebutuhan makanan, garis kemiskinan berdasarkan kebutuhan nonmakanan, dan garis kemiskinan total. Dari aspek kelengkapan data, terdapat kekosongan nilai garis kemiskinan pada beberapa data dengan kriteria berikut.

1. Data garis kemiskinan total pada seluruh provinsi pada tahun 2013 dan 2014.
2. Sebagian data garis kemiskinan pada provinsi DKI Jakarta pada kategori nonmakanan dan total perdesaan

Secara kualitas, terdapat ambiguitas data terkait kategori daerah, dimana nilai kategori daerah terbagi menjadi “PEDESAAN”, “PERKOTAAN”, dan “PERDESAANPERKOTAAN”. Tidak ada penjelasan lebih jelas tentang kategori “PERDESAANPERKOTAAN”. Selain itu, tidak ada penjelasan tentang perbedaan jenis dan harus dilakukan pencarian informasi tentang informasi garis kemiskinan total, yang merupakan hasil dari penjumlahan nilai garis kemiskinan makanan dan garis kemiskinan nonmakanan. Setelah melakukan evaluasi manual, ditemukan bahwa terdapat beberapa data garis kemiskinan total yang memiliki selisih sedikit apabila dibandingkan dengan penjumlahan garis kemiskinan makanan dan garis kemiskinan nonmakanan. Pada dataset ini tidak terdapat data duplikat.

Berikut adalah tabel tipe data pada dataset garis kemiskinan.

#### Tabel II.3 Tipe Data dan Deskripsi Tambahan dalam Dataset GK

| **Nama Kolom** | **Tipe Data** | **Deskripsi** |
| --- | --- | --- |
| Provinsi | Kategorikal nominal | 34 provinsi unik Indonesia sebelum tahun 2022. |
| Jenis Pengeluaran | Kategorikal nominal | 3 kategori pengeluaran unik yang dibagi ke dalam makanan, nonmakanan, dan total. |
| Daerah Tempat Tinggal | Kategorikal nominal | 3 kategori daerah unik yang dibagi ke dalam perkotaan, perdesaan, dan keduanya. |
| Tahun | Kuantitatif interval | Tahun di saat data diperoleh dalam rentang 2013 hingga 2022. |
| Periode | Kategorikal ordinal | 2 periode unik dalam satu tahun. |
| GK | Kuantitatif rasio | Pendapatan minimum yang diperlukan untuk memenuhi standar hidup yang cukup dalam satuan Rupiah. |

### Dataset Pengeluaran per Kapita (PPK)

Dataset ini merepresentasikan pengeluaran per kapita untuk setiap provinsi dalam rentang tahun 2007 hingga 2022. Dari hasil evaluasi, terdapat kekosongan data pengeluaran per kapita pada beberapa provinsi, yaitu provinsi DKI Jakarta pada daerah pedesaan serta provinsi Kalimantan Utara yang tidak memiliki data pengeluaran per kapita sebelum tahun 2014. Berikut adalah tabel mengenai tipe data serta deskripsi tambahan mengenai kolom dalam pengeluaran per kapita. Pada dataset ini tidak terdapat data duplikat.

#### Tabel II.4 Tipe Data dan Deskripsi Tambahan dalam Dataset PPK

| **Nama Kolom** | **Tipe Data** | **Deskripsi** |
| --- | --- | --- |
| Provinsi | Kategorikal nominal | 34 provinsi unik Indonesia sebelum tahun 2022. |
| Jenis Pengeluaran | Kategorikal nominal | 3 kategori pengeluaran unik yang dibagi ke dalam makanan, nonmakanan, dan total. |
| Daerah Tempat Tinggal | Kategorikal nominal | 3 kategori daerah unik yang dibagi ke dalam perkotaan, perdesaan, dan keduanya. |
| Tahun | Kuantitatif interval | Tahun di saat data diperoleh dalam rentang 2007 hingga 2022. |
| PPK | Kuantitatif rasio | Pengeluaran per kapita pada suatu daerah dalam satuan Rupiah. |

## Transforming for Quality

Pada tahapan transformasi untuk kualitas, kami melakukan penghilangan data kotor dengan menghapus data yang tidak bersih dan evaluasi apakah terdapat data duplikat. Kami menentukan ini untuk mempermudah pengolahan data dan penarikan analisis berdasarkan jenis atau daerah tertentu selanjutnya. Untuk perhitungan kategori jenis total, akan dilakukan perhitungan secara manual menggunakan data jenis makanan dan nonmakanan mengingat terdapat beberapa selisih yang tipis antara perhitungan data total dan jumlah data jenis makanan dan nonmakanan. Kami juga memutuskan untuk menghapus data garis kemiskinan dan pengeluaran per kapita pada provinsi DKI Jakarta pada daerah pedesaan dikarenakan tidak adanya nilai pencatatan pada kedua data tersebut. Data dengan provinsi ‘Indonesia’ juga tidak akan dimasukkan karena data tersebut mengacu ke akumulasi informasi ke satu data yang mewakili negara.

## Transforming for Analysis

Untuk mempermudah pemrosesan data garis kemiskinan, kami memutuskan untuk hanya menggunakan data pada bulan September. Hal ini dikarenakan data pada bulan September yang mendekati nilai garis kemiskinan akhir pada suatu tahun, sehingga lebih baik menyederhanakan data dengan menggunakan data pada bulan September.

Salah satu transformasi yang dilakukan adalah menambahkan informasi lain yang diperoleh melalui pemrosesan data dari tabel. Misalnya, di data tabular mengenai upah per jam, kami menambahkan informasi upah bulanan yang merupakan estimasi perolehan bulanan menggunakan harga upah per jam. Selain itu, kami menambahkan informasi apakah upah bulanan hasil estimasi lebih tinggi dibandingkan UMP yang tercatat untuk mengetahui jumlah provinsi dengan estimasi pendapatan rerata di bawah UMP. Terdapat peluang untuk melakukan agregasi melalui perhitungan rerata untuk data garis kemiskinan dan pengeluaran per kapita dikarenakan tidak adanya informasi tentang perbedaan daerah, sehingga agregasi melalui kalkulasi rata-rata dapat merepresentasikan data secara lebih keseluruhan. Apabila diperlukan perbandingan antar daerah atau dapat jenis, data yang dimiliki dapat dipisah berdasarkan kategori untuk melakukan visualisasi per kategori.

Selain itu, kami melakukan *filtered resolution* terhadap data dengan menggunakan data di atas tahun 2015. Diketahui bahwa data upah tercatat mulai tahun 2015, data UMP tercatat mulai tahun 2002, data garis kemiskinan tercatat mulai tahun 2013, dan data pengeluaran per kapita tercatat mulai tahun 2007. Hal ini menandakan bahwa kelengkapan data untuk analisis keseluruhan dapat dilakukan pada data yang direkam mulai tahun 2015. Selain itu, terdapat kekosongan data garis kemiskinan pada periode 2013 - 2014 serta kekosongan data pengeluaran per kapita pada provinsi Kalimantan Utara pada periode 2007 - 2013. Oleh karena itu, kami menetapkan untuk hanya menggunakan data di atas tahun 2015 untuk analisis selanjutnya dan membuang data di bawah tahun 2015.

## Consolidating

Pada tahap ini, dilakukan evaluasi terhadap data yang telah didapatkan. Adapun untuk memenuhi tujuan yang telah ditetapkan, diperlukan suatu tambahan data terkait kondisi lapangan pekerjaan di Indonesia. Data tersebut berfungsi sebagai visualisasi pendukung dari visualisasi utama yang akan dikembangkan. Visualisasi kondisi lapangan pekerjaan dapat pula dijadikan sebagai tambahan pertimbangan bagi pengguna untuk memahami kondisi serta kesempatan pekerja di seluruh Indonesia.

# QUESTION FORMULATION

## Editorial Focus

Fokus editorial pada visualisasi data ini adalah menunjukkan bagaimana kesenjangan tingkat kesejahteraan pekerja antar provinsi di Indonesia dan dorongan untuk menimbang ulang pemilihan daerah untuk bekerja. Dengan fungsionalitas visualisasi data berupa *explanatory*, fokus editorial yang dibuat dapat berisi serangkaian penjelasan tambahan pada beberapa titik di visualisasi agar *insight* yang didapatkan lebih menyeluruh.

Tipe fokus editorial yang dipilih adalah *explanatory*. Melalui visualisasi ini, kami bertujuan untuk menyampaikan wawasan dengan narasi spesifik mengenai tren UMP, garis kemiskinan, serta pengeluaran per kapita di berbagai provinsi di Indonesia. Oleh karena itu, tipe fokus *explanatory* dipilih karena sifatnya yang bersifat menjelaskan dan menyampaikan wawasan-wawasan tertentu dari data yang diperoleh.

Karakteristik fokus editorial yang dipilih adalah perbandingan dan proporsi. Hal ini disebabkan oleh sifat visualisasi yang membandingkan berbagai variabel kuantitatif diskrit dalam satu kategori yang sama. Selain itu, karakteristik lain yang dipilih berupa tren dan pola, karena visualisasi ini juga bertujuan untuk memvisualisasikan perubahan nilai seiring berjalannya waktu.

## Reasoning

Penalaran yang digunakan pada visualisasi data adalah penalaran deduktif. Hal ini dikarenakan adanya suatu pra asumsi mengenai tingkat kesejahteraan pekerja di Indonesia yang telah ditentukan sebelumnya. Visualisasi data dikembangkan dengan menimbang dan memvalidasi pra asumsi yang telah ditentukan sebelumnya sehingga didapatkan suatu kunci atau *insight* pada *scope* tersebut. Pra Asumsi yang dipilih dalam penalaran ini berkaitan dengan kesenjangan kesejahteraan pekerja pada berbagai daerah dan ketercukupan pendapatan pekerja dibandingkan terhadap pengeluarannya.

## Narrative

### Genre

Visualisasi yang dibuat akan menggunakan *genre* *annotated chart*. Genre ini menekankan pembuatan berbagai visualisasi dengan adanya anotasi mengenai wawasan yang diperoleh melalui properti-properti visual. Hal ini selaras dengan tujuan visualisasi kami untuk mempermudah masyarakat awam memperoleh wawasan melalui visualisasi data. Dengan adanya visualisasi statistik, diharapkan hal-hal penting dalam data dapat dianotasi dan diolah menjadi wawasan.

### Visual Narrative Tactics

Taktik narasi visual yang dibuat akan menggunakan *highlighting*. Penggunaan taktik narasi visual tersebut dapat dilakukan melalui penggunaan warna yang lebih mencolok pada penekanan visualisasi. *Highlighting* digunakan agar pembaca dapat dengan mudah melihat sekumpulan poin-poin penting yang ditunjukkan pada visualisasi data. Hal ini juga dapat membantu pengguna untuk menemukan suatu *insight* pada data dengan lebih efisien dan efektif.

### Narrative Structure Tactics

Taktik struktur narasi yang kami pilih adalah *messaging*. *Messaging* merujuk ke bagaimana sebuah visualisasi dapat mengkomunikasikan observasi dan komentar ke pengguna. Komunikasi ini dapat dilakukan melalui teks seperti *headline* atau anotasi, atau deskripsi-deskripsi tertentu seperti rangkuman atau artikel. Hal ini akan kami gunakan untuk menyampaikan informasi yang diperoleh dari sebuah visualisasi ke para pengguna yang akan membaca.

# 

# REFERENSI

Yakhamid, R.Y. (2023) *Kesejahteraan Pekerja Indonesia*, *Kaggle*. Available at: https://www.kaggle.com/datasets/rezkyyayang/pekerja-sejahtera (Accessed: 02 March 2024).