**GAJIMU.COM**

**Recruitment Data Analyst**

**Name : Ferdiansyah Ersatiyo**

Soal:

1. Berikan penjelasan berdasarkan pengalaman/pendapat Anda untuk pertanyaan dibawah ini:

a. Apa langkah-langkah yang diperlukan untuk membersihkan data yang kotor/tidak lengkap dan data duplikat?

b. Bagaimanakah peranan Data Analyst dalam dalam konteks upah, kondisi kerja, atau masalah-masalah seputar tenaga kerja?

c. Bagaimana Anda berkomunikasi hasil analisis data kepada organisasi terkait yang mungkin tidak memiliki latar belakang teknis?

2. Buatlah visualisasi dari data berikut menggunakan Google Looker Studio atau Tableau:

[Survey Upah Pekerja Pabrik dan Upah Minimum Kota:](https://docs.google.com/spreadsheets/d/1wMR2w-4MSsehI6X6GjfTITG8K8SwhYjdDsclNUzG3cA/edit#gid=1387715644)

Jawaban :

1. A. Dalam melakukan analisis data diperlukan data clean atau data yang sudah dibersihkan, Cara untuk membersihkan data menurut saya yang paling utama adalah cara kita mendefiniskan data bersih seperti apa? Perlakuan cleaning data pada setiap data bisa berbeda beda dan tidak selalu sama, namun ada beberapa hal yang paling umum dilakukan yaitu checking duplicated data dan missing values. Dalam menghandle duplicated value kita tidak bisa langsung mendelete salah satu rows data agar tidak adanya duplicated data, untuk contoh data pasien rumah sakit pasti memiliki nama pasien dan jenis penyakit yang dialami, bisa saja pasien mengalami sakit yang sama lebih dari 1 kali sehingga terjadi duplikat data namun apabila data tersebut merupakan duplikat yang valid kita dapat melakukan beberapa cara seperti
   1. Delete duplicate data
   2. Keep first data delete rest duplicate
   3. Rata2 nilai duplicate
   4. Mark duplicate

Untuk menghandle missing values yang harus dipahami adalah tipe dari missing value itu sendiri ada 3 Tipe Missing values

1. MCAR (Missing Completely at Random)

MCAR ini adalah tipe missing values yang benar-benar tidak ada atau hilang pada saat proses pengumpulan data sehingga kita tidak dapat melakukan pengisian apapun pada missing value tipe ini dan dapat melakukan delete data namun apabila banyak data yang memiliki MCAR pada sebuah dataset dan dapat merubah hasil secara signifikan sebaiknya tidak dilakukan delete dan didiamkan saja.

1. MAR (Missing at Random)

MAR sendiri adalah tipe data yang bisa dilakukan imputasi dengan cara mengobservasi data yang berhubungan pada data yang hilang tersebut untuk handling MAR dapat melakukan imputasi data menggunakan nilai modus atau mean dari sebuah dataset

1. MNAR (Missing Not at Random)

MNAR adalah tipe data yang hilang atau sengaja tidak berisi karena bersifat seperti informasi sensitif seorang responden. Untuk menghandling MNAR lebih sulit karena bisa jadi sebuah missing values berkaitan dengan data lainnya.

B. Data Analyst memiliki peran penting dalam konteks upah, kondisi kerja, dan masalah-masalah seputar tenaga kerja, mengapa demikian? Karena data analyst memiliki peran untuk menginterpretasi data yang berantakan menjadi sebuah informasi yang dapat dengan mudah dipahami oleh masyarakat umum. Seperti menghiglight tentang kenaikan gaji UMP yang baru baru ini terjadi, mengkritisi kondisi kerja yang terjadi dipabrik pabrik garmen dan lainnya dan dari hasil analisis data dapat menghiglight perusahaan yang tidak mengikuti aturan yang berlaku.

C. Untuk berkomunikasi dengan organisasi yang tidak memiliki latar belakang teknis sebenarnya cukup mudah,Bagaimana caranya? Manusia sejatinya adalah mahluk visual,dengan membuat visualisasi dari data yang sudah diolah dengan bentuk visualisasi yang sederhana dan mudah dipahami akan memperlancar proses komunikasi yang terjadi diantara kedua belah pihak sehingga dapat terjadi kesepakatan yang saling menguntungkan dari kedua belah pihak.

2. Visualisasi Data Survey Upah Pekerja Pabrik dan Upah Minimum Kota

1. Perencanaan
   1. Berapa banyak pekerja yang dibayar sama dengan atau di atas batas upah minimum dan berapa pekerja yang dibayar di bawah batas upah minimum?
   2. Berapa banyak pabrik yang membayar pekerja sama dengan atau di atas batas upah minimum dan berapa pabrik yang membayar pekerja di bawah batas upah minimum?
   3. Apakah ada kesenjangan gaji berdasarkan gender pekerja
   4. Apakah ada kesenjangan gaji antara usia muda dan usia tua
2. Proses

* Checking Missing Values dan duplikat data pada google sheets
* Upload Google Sheets ke Bigquery
* Change nama column Region menjadi Daerah\_Region dari sheets MW\_Rate agar tidak terjadi confusion karena Colum REGION ada di 2 sheets dengan nama yang sama kemudian save as view
* Query Gaji ke dalam new colum dengan perhitungan periode pembayaran \* jumlah pembayaran sesuai ketentuan

1 : Upah dibayarkan sebulan sekali

2 : Upah dibayarkan 4 minggu sekali

3 : Upah dibayarkan 2 minggu sekali

4 : Upah dibayarkan seminggu sekali

6 : Upah dibayarkan per jam.

* Join 3 table yang ada sehingga mendapatkan gaji umr di tiap kota kabupaten , dan nama perushaan
* Lakukan pengelompokan apakah termasuk ke dalam gaji diatas umr, gaji umr atau gaji dibawah umr dengan menggunakan CTE Table
* Lakukan pengelompokan berdasarkan usia
* Berikut Querry yang dilakukan

# Checking Null Values

# im doing the checking in google sheets because more easy to do it there.

#Change Column Name for avoiding confusion when joining the table

SELECT

REGION as Daerah\_Region,

\*

FROM `propane-girder-388207.GajimurRecruitmentDataAnalyst.MW\_Rate`;

# save this as a view

#CLEANING THE DATA

WITH GAJI\_CTE AS(

SELECT

\*,

CASE

WHEN gaji.On\_what\_payment\_period\_was\_your\_last\_wage\_based\_ = 6 THEN (gaji.Wage\_per\_\_Period)\*160

WHEN gaji.On\_what\_payment\_period\_was\_your\_last\_wage\_based\_ = 4 THEN (gaji.Wage\_per\_\_Period)\*4

WHEN gaji.On\_what\_payment\_period\_was\_your\_last\_wage\_based\_ = 3 THEN (gaji.Wage\_per\_\_Period)\*2

ELSE (Wage\_per\_\_Period)\*1

END as Total\_gaji\_bulan

FROM`propane-girder-388207.GajimurRecruitmentDataAnalyst.MW\_Cleaning` as gaji

JOIN `propane-girder-388207.GajimurRecruitmentDataAnalyst.Clean\_UMR` as UMR

ON gaji.REGION = UMR.Daerah\_Region

JOIN `propane-girder-388207.GajimurRecruitmentDataAnalyst.Comp\_API` as comp

ON gaji.COMPANY\_API2\_PROJECT = comp.COMPANY\_API\_CODE

)

SELECT

CASENUM,

KEY,

SEX,

AGE,

COMPANY\_NAME,

REGION,

Applicable\_MW\_Rate,

Total\_gaji\_bulan,

CASE

WHEN Total\_gaji\_bulan > Applicable\_MW\_Rate THEN 'Diatas UMR'

WHEN Total\_gaji\_bulan = Applicable\_MW\_Rate THEN 'GAJI UMR'

WHEN Total\_gaji\_bulan < Applicable\_MW\_Rate THEN 'Dibawah UMR'

END AS Kategori,

CASE

WHEN AGE < 18 THEN 'Underage'

WHEN AGE >= 18 AND AGE < 29 THEN 'Young Adult'

WHEN AGE >= 29 AND AGE < 40 THEN 'Adult'

WHEN AGE >= 40 AND AGE < 51 THEN 'Middle Age'

ELSE 'Senior'

END AS Age\_Category

FROM GAJI\_CTE;

# Berapa banyak Pekerja yang menerima gaji diatas atau sama dengan UMR dan berapa banyak Pekerja yang menerima gaji dibawah UMR

WITH Gaji\_cat AS(

SELECT

CASENUM,

SEX,

CASE

WHEN Kategori = 'Diatas UMR' THEN 'diatas atau sama dengan UMR'

WHEN Kategori = 'GAJI UMR' THEN 'diatas atau sama dengan UMR'

ELSE 'dibawah UMR'

END AS Category\_gaji

FROM `propane-girder-388207.GajimurRecruitmentDataAnalyst.Clean\_Data`

)

SELECT

Category\_gaji,

COUNT(Category\_gaji) as total\_category\_gaji

FROM Gaji\_cat

GROUP BY 1;

#Berapa banyak pabrik yang membayar pekerja sama dengan atau di atas batas upah minimum dan berapa pabrik yang membayar pekerja di bawah batas upah minimum?

SELECT

DISTINCT(COMPANY\_NAME),

COUNT(Kategori) Pegawai\_dibawah\_UMR

FROM `propane-girder-388207.GajimurRecruitmentDataAnalyst.Clean\_Data`

WHERE Kategori = 'Dibawah UMR' AND COMPANY\_NAME IS NOT NULL

GROUP BY 1

ORDER BY 2 DESC

#Melihat Daerah dimana masih banyak pekerja yang dibayar dibawah UMR

WITH Gaji\_cat AS(

SELECT

REGION,

CASE

WHEN Kategori = 'Diatas UMR' THEN 'diatas atau sama dengan UMR'

WHEN Kategori = 'GAJI UMR' THEN 'diatas atau sama dengan UMR'

ELSE 'dibawah UMR'

END AS Category\_gaji

FROM `propane-girder-388207.GajimurRecruitmentDataAnalyst.Clean\_Data`

)

SELECT

REGION,

COUNT(CASE WHEN Category\_gaji = 'dibawah UMR' THEN 1 END) AS dibawah\_UMR,

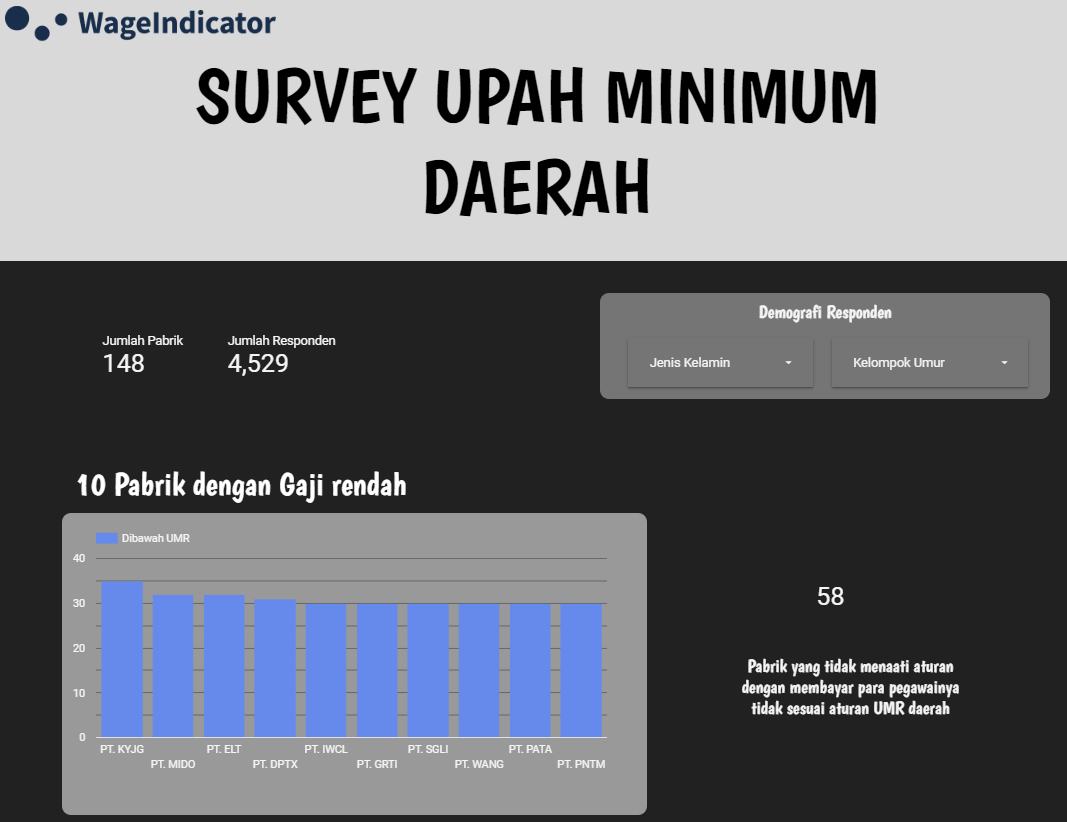
COUNT(CASE WHEN Category\_gaji = 'diatas atau sama dengan UMR' THEN 1 END) AS diatas\_atau\_sama\_dengan\_UMR

FROM Gaji\_cat

GROUP BY REGION

ORDER BY 2 DESC

1. Hasil



DASHBOARD LINK :

<https://lookerstudio.google.com/reporting/3fd00a6a-18ef-4ef2-8c97-4df061fa99ed>

Pertanyaan untuk dijawab:

1. Berapa banyak pekerja yang dibayar sama dengan atau di atas batas upah minimum

dan berapa pekerja yang dibayar di bawah batas upah minimum?

2. Berapa banyak pabrik yang membayar pekerja sama dengan atau di atas batas upah

minimum dan berapa pabrik yang membayar pekerja di bawah batas upah

minimum?

3. Insight apa saja yang didapatkan dari data diatas?

Jawab

1. Sebanyak 3717 orang menerima gaji setara umr atau diatas umr sedangkan sebanyak 842 orang masih menerima gaji dibawah standar umr yang ditetapkan pada tiap daerah.
2. Sebanyak 85 Pabrik masih memberikan gaji dibawah standar yang ditentukan pada tiap daerah
3. Kabupaten purwarkarta menjadi daerah yang paling banyak menggaji karyawan dibawah standar umr yang ditentukan.
4. Pada setiap kelompok umur terdapat pegawai yang masih menerima gaji dibawah umr sehingga dapat dikatakan bahwa umur tidak berkaitan dengan gaji yang diterima.