

Data

- Dataset merupakan data delivery task selama 10 hari dari '2022-11-01' sampai '2022-11-10' dengan format JSON
- Total data yaitu sebanyak 8334 rows
- Terdapat 18 fitur yang ada di dataset, deskripsi dari fitur tersebut telah diberikan di [Github link](#)

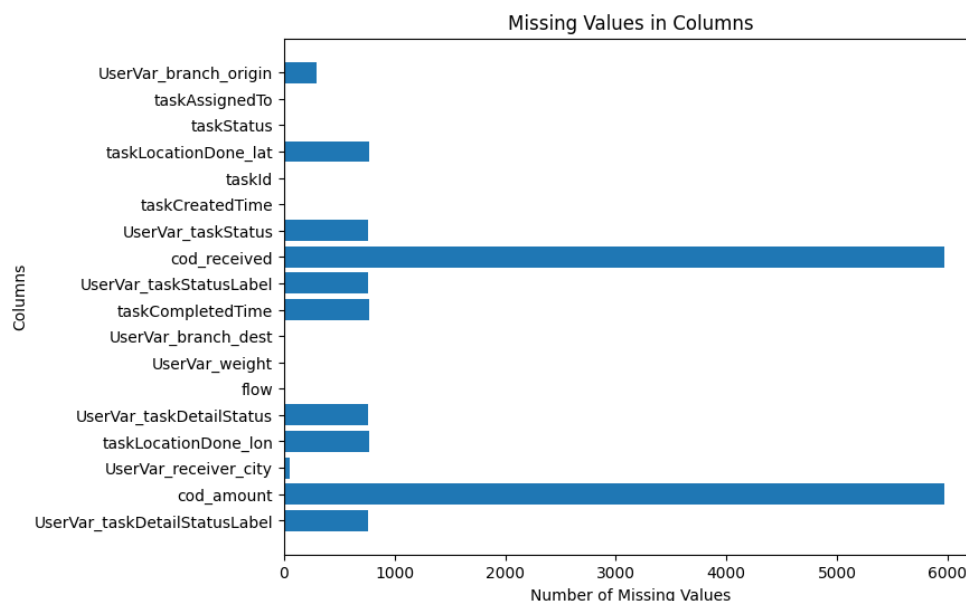
Field	Description
taskId	Unique identifier for the task that generated by system.
taskCreatedTime	Time at when the task was created.
taskCompletedTime	Time at when the task was completed.
taskAssignedTo	Worker that doing the task.
taskLocationDone	Coordinate of where the task was completed.
flow	Flow or type of the task.
cod	Contains data for the COD system.
cod.amount	Amount of money from COD.
cod.received	COD has been received or not.
UserVar	Contains more specified data, in this case the 'UserVar' is about delivery task data.
UserVar.taskStatus	Delivery status code.
UserVar.taskStatusLabel	Delivery status label. <ul style="list-style-type: none">• Success = package is not delivered to recipient• Failed = package is not delivered to recipient
UserVar.taskDetailStatus	Detailed delivery status code.
UserVar.taskDetailStatusLabel	Detailed delivery status label.
UserVar.branch_origin	Branch code of the origin.

UserVar.branch_dest	Branch code of the destination.
UserVar.weight	Weight of the package.
UserVar.receiver_city	City of the destination
taskStatus	Task status <ul style="list-style-type: none"> • Done = task is already done • Ongoing = task is still ongoing

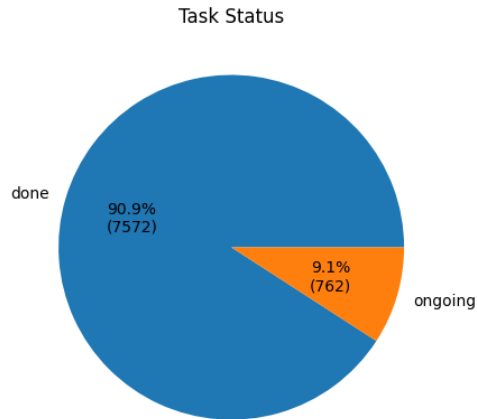
Analisa Data

Pembersihan Data

- Mengecek apakah ada missing values pada data



Terdapat missing values pada beberapa kolom dan banyak ditemukan di kolom **cod**, dikarenakan tidak semua delivery task menggunakan opsi Cash on Delivery (COD). Kolom **UserVar_taskDetailStatus**, **UserVar_taskDetailStatusLabel**, **UserVar_taskStatusLabel**, **UserVar_taskStatus** memiliki 762 missing values dan kolom **taskCompletedTime**, **taskLocationDone_lon**, **taskLocationDone_lat** memiliki 768 missing values.



Jika dilihat dari unique values kolom **taskStatus**, terdapat 762 values yang merupakan 'ongoing'. Dan 762 missing values yang ada di kolom **UserVar_taskDetailStatus**, **UserVar_taskDetailStatusLabel**, **UserVar_taskStatusLabel**, **UserVar_taskStatus**, **taskCompletedTime**, **taskLocationDone_lon**, **taskLocationDone_lat** memiliki status 'ongoing'. Dapat disimpulkan bahwa data pada kolom-kolom tersebut belum terisi karena status task tersebut yaitu 'ongoing'.

```
check_taskDetailStatus = df[df['UserVar_taskStatusLabel'].isna()]
check_taskDetailStatus.taskStatus.value_counts()
```

```
ongoing    762
Name: taskStatus, dtype: int64
```

```
check_taskStatusLabel = df[df['UserVar_taskStatusLabel'].isna()]
check_taskStatusLabel.taskStatus.value_counts()
```

```
ongoing    762
Name: taskStatus, dtype: int64
```

```
check_taskCompletedTime = df[df['taskCompletedTime'].isna()]
check_taskCompletedTime.taskStatus.value_counts()
```

```
ongoing    762
done        6
Name: taskStatus, dtype: int64
```

6 missing values yang terdapat pada kolom **taskCompletedTime**, **taskLocationDone_lon**, **taskLocationDone_lat** memiliki status 'done' dan **UserVar_taskStatusLabel** = 'success' yang berarti paket telah diterima oleh penerima. 6 task tersebut juga dikerjakan oleh worker(taskAssignedTo) yang sama dan tujuan yang sama yaitu BABELAN,CIKARANG, sehingga dapat diasumsikan task tersebut telah berhasil diselesaikan namun untuk kolom **taskCompletedTime**, **taskLocationDone_lon**, **taskLocationDone_lat** tidak terisi. Begitu juga dengan kolom lain seperti **UserVar_branch_origin** dan **UserVar_receiver_city** yang masing-masing

memiliki 293 dan 52 missing values, dapat diasumsikan data tersebut hilang (tidak memiliki pattern tertentu).

- Mengubah format untuk kolom **taskCreatedTime** dan **taskCompletedTime** menjadi datetime. Timestamp yang ada pada kedua kolom tersebut memiliki UTC yang berbeda misalkan +0700 dan +0800 yang menandakan bahwa timestamp tersebut bisa tercatat sebagai WIB atau WITA atau WIT. Agar timestamp dapat dibaca dengan lebih mudah, maka semua value akan diubah menjadi UTC+7. Sebelum diubah menjadi datetime, missing values yang ada di **taskCompletedTime** akan diisi dengan value dari **taskCreatedTime** sehingga durasi task adalah 0.

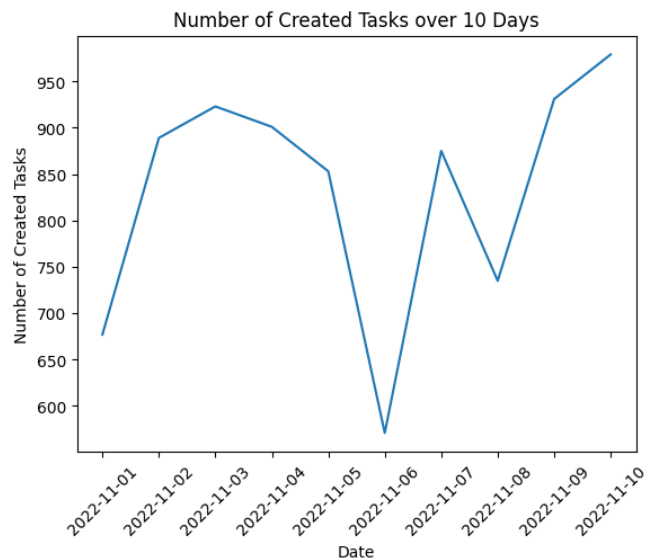
Feature Engineering

Untuk menganalisa data lebih jauh, maka akan dibentuk fitur-fitur baru dari data yang sudah ada:

- taskDuration = durasi untuk tiap task dalam menit
- createdDow dan completedDow = hari dalam minggu (0 untuk hari minggu)
- createdHour dan completedHour = jam task tersebut

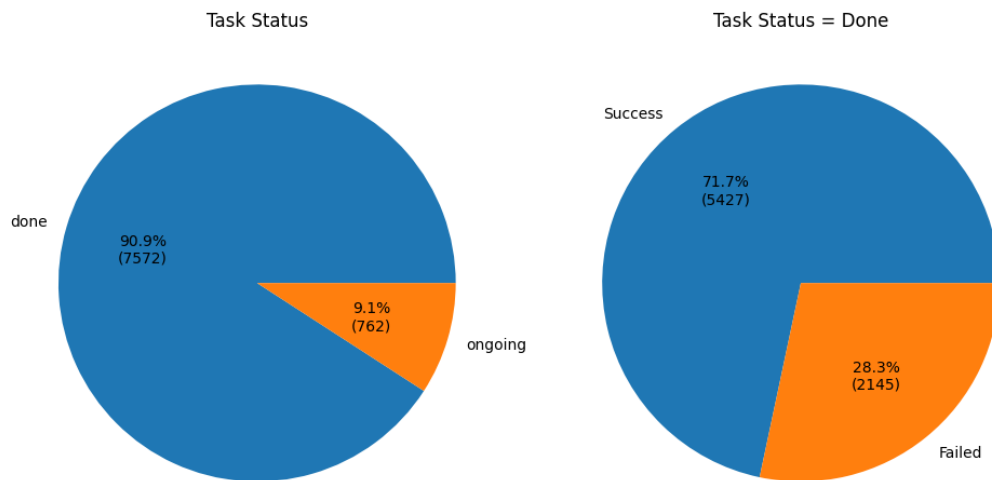
Insight

Jumlah Task



Jumlah task untuk tiap harinya berbeda, jumlah task paling banyak adalah 979 task di tanggal 10 November 2022 (hari kamis) dan jumlah task paling sedikit di tanggal 6 November 2022 (hari minggu) yaitu sebanyak 571 task.

Task Status



Dari keseluruhan data dalam periode 10 hari, terdapat 90.9% task yang telah selesai (*done*) dan 9.1% task yang masih dijalankan (*ongoing*). Namun dari task yang telah selesai (*done*), terdapat hanya 71.7% task yang berhasil diterima oleh penerima (*success*) dan 28.3% task tidak berhasil (*failed*).

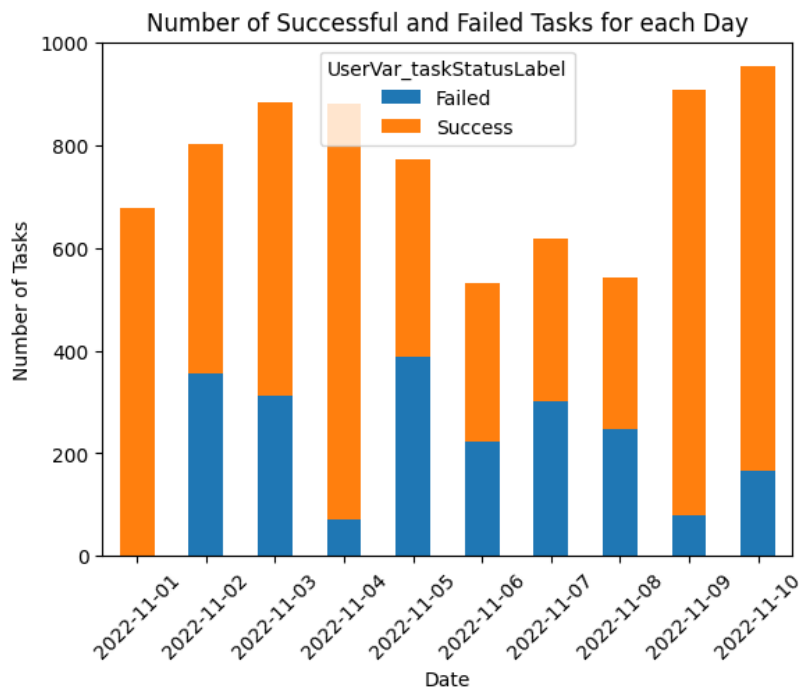
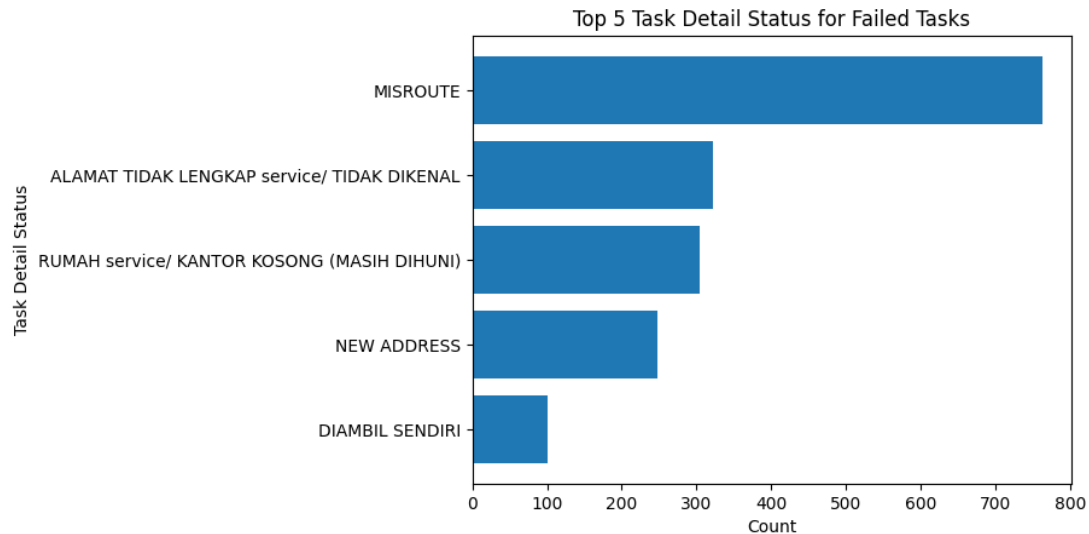


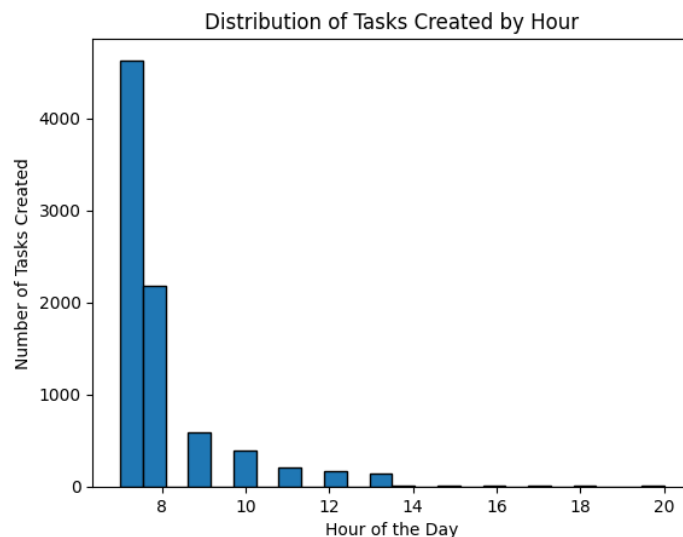
Chart di atas merupakan distribusi Done task dengan status 'Success' dan 'Failed' untuk tiap harinya dalam periode 10 hari. Dapat dilihat pada tanggal 1 November 2022 semua task berhasil diterima oleh penerima, dan pada tanggal 5 November 2022 lebih dari 50% task memiliki status 'Failed'.

Untuk mengetahui alasan task tidak berhasil (dalam periode 10 hari) adalah dengan melihat fitur **UserVar_taskDetailStatusLabel**. Fitur tersebut memberikan detail siapa penerima paket atau dapat berupa informasi lain. Berikut ini merupakan 5 detail pengiriman atau alasan terbanyak mengapa task tersebut gagal.

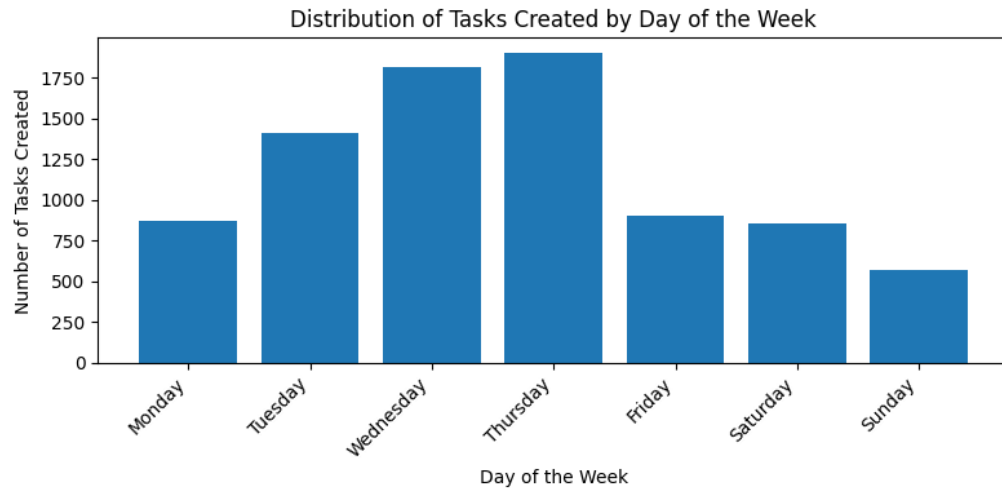


Alasan terbanyak adalah *misroute* dan 4 alasan lainnya bisa diasumsikan merupakan kesalahan dari penerima seperti tidak memberikan alamat lengkap atau tidak memberikan alamat yang baru. *Misroute* atau salah rute memiliki jumlah yang banyak yaitu 763 rows dan bisa diasumsikan merupakan kesalahan dari kurir, sehingga dapat diharapkan untuk kedepannya *misroute* ini dapat diatasi.

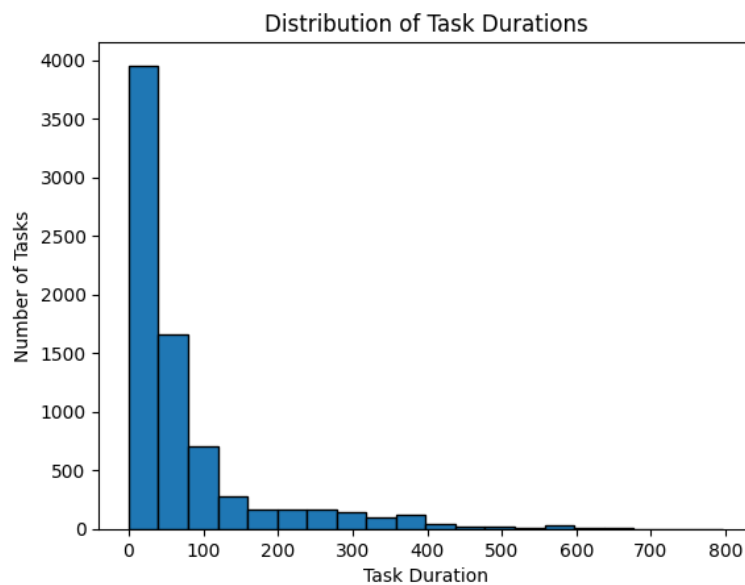
Waktu



Dalam periode 10 hari, jumlah task yang paling banyak dibuat yaitu pada pukul 7 pagi. Rata-rata dan median dari distribusi task yang dibuat juga pada pukul 7 pagi.



Untuk distribusi task dalam bentuk hari per minggu, jumlah task yang paling banyak dibuat yaitu pada hari kamis. Sedangkan untuk rata-ratanya adalah pada hari rabu dan mediannya adalah hari kamis.



Dari task yang memiliki status 'done', kita bisa menghitung durasi task tersebut. Chart di atas merupakan distribusi durasi task, dapat dilihat bahwa ~20% task dapat diselesaikan (menjadi 'done') dalam waktu 5 menit. Sedangkan untuk rata-ratanya yaitu ~1 jam dan mediannya adalah ~37 menit.

Pekerja/Kurir

Jumlah pekerja/kurir yang diberikan task dalam periode waktu 10 hari sebanyak 2787 orang.

Untuk keseluruhan data

- Dalam periode waktu 10 hari : rata-rata 1 pekerja mengerjakan 2-3 task, dengan jumlah task paling sedikit yaitu 1 task dan jumlah task yang paling banyak adalah 103. Pekerja yang memiliki task terbanyak adalah **gutturalLion9**
- Dalam 1 hari : rata-rata 1 pekerja bisa mengerjakan 2-3 task, dengan jumlah task paling sedikit yaitu 1 task dan jumlah task yang paling banyak adalah 83. Pekerja yang memiliki task terbanyak adalah **gloomyLlama0**

Untuk data dengan taskStatus = 'done', analisa success rate atau berapa task yang berhasil diterima oleh penerima (UserVar_taskStatusLabel = 'Success') untuk tiap pekerja

- Dalam periode waktu 10 hari : rata-rata 1 pekerja memiliki 95% success rate atau 95% tasknya memiliki status 'Success', dan dengan success rate paling rendah yaitu 2% dan paling tinggi adalah 100% yang berarti semua task memiliki status 'Success'. Terdapat 2020 pekerja yang memiliki success rate yaitu 100%.
- Dalam 1 hari : rata-rata 1 pekerja memiliki 98% success rate atau 98% tasknya memiliki status 'Success', dan dengan success rate paling rendah yaitu 3% dan paling tinggi adalah 100% yang berarti semua task memiliki status 'Success'. Terdapat 2860 pekerja yang memiliki success rate yaitu 100%.

Branch Origin dan Destination

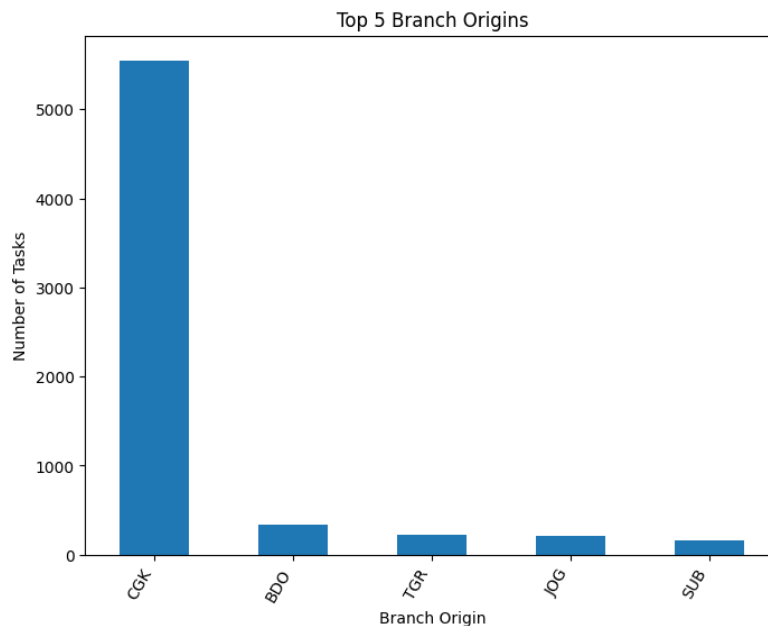
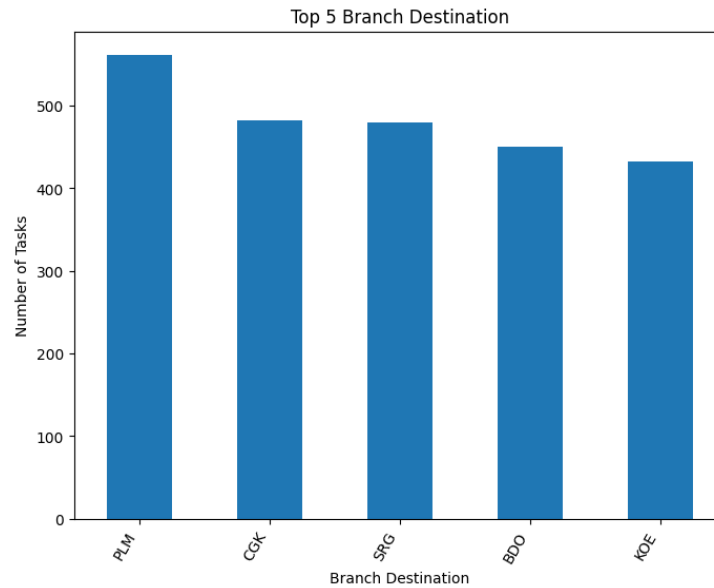
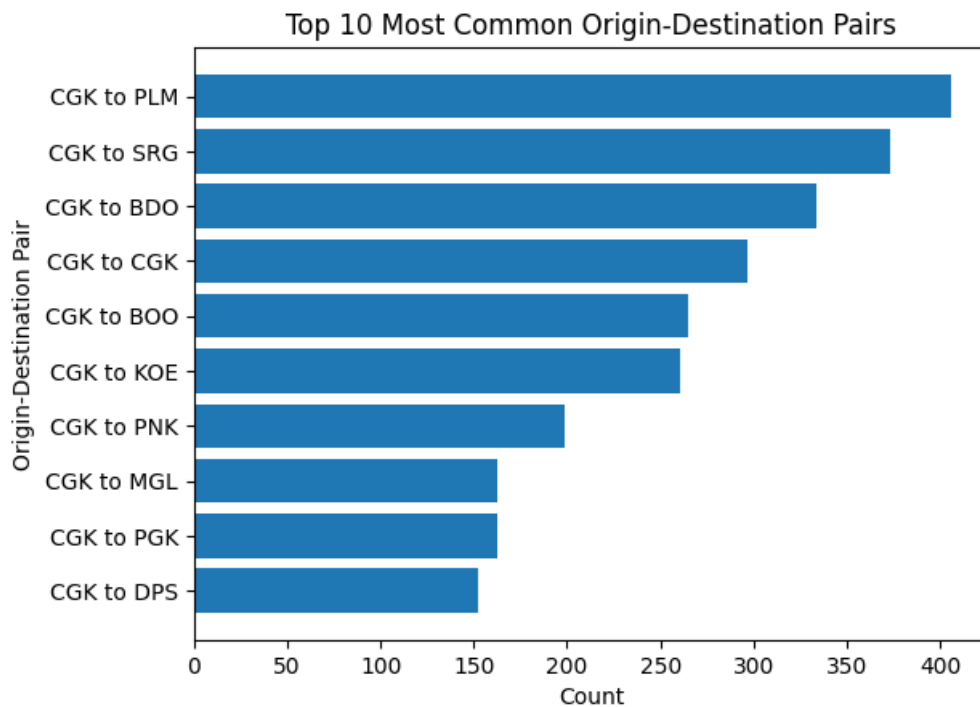


Chart di atas merupakan 5 branch origin terbanyak dalam periode 10 hari. Dapat dilihat bahwa jumlah task yang berasal dari CGK mencapai lebih dari 5000, dan 4 branch lainnya hanya mencapai angka di bawah 500 task. Analisa ini membuktikan bahwa banyak pengguna jasa berasal dari Jakarta dan sekitarnya.



Sedangkan untuk branch destination, dapat dilihat bahwa jumlah task dengan tujuan PLM mencapai lebih dari 500, dan 4 branch lainnya hanya mencapai angka di bawah 500 task. Analisa ini membuktikan bahwa banyak penerima berasal dari Palembang dan sekitarnya.



Analisa lebih lanjut yaitu melihat distribusi pasangan branch origin dengan branch destination. Bar chart di atas menunjukkan 10 branch origin-branch destination yang paling banyak dalam periode 10 hari. Pair CGK-PLM adalah yang paling banyak hingga mencapai angka 400. 9 pair lainnya juga masih berasal dari branch CGK yang menandakan bahwa pengguna jasa banyak berasal dari Jakarta dan sekitarnya.

Github Link :

<https://github.com/jihannbl/Mileapp-Test>

Jupyter Notebook :

<https://colab.research.google.com/drive/1tNL3olypgWB5nQ0lhTRKNew6TIQI202a?usp=sharing>