

Data Analysis for Finance :

Performa Cabang

Project from DQLab
Mentor : Farkhan Novianto, Head of Data
Analytics AMARTHA



Bintang Apriansyah Tambunan

ABOUT PROJECT

DQLab Finance merupakan perusahaan *finance* yang konsisten menyalurkan pembiayaan untuk masyarakat dan sudah memiliki banyak cabang.



TUGAS SEBAGAI DATA ANALYST

Menganalisis bagaimana performa cabang pada bulan Mei 2020

Langkah yang dilakukan :

1. Memfilter data untuk bulan Mei 2020
2. Membuat *summary* per cabang untuk melihat data 5 cabang terbaik dan terburuk
3. Periksa umur cabang dan performa bulan Mei 2020
4. Mencari cabang terburuk untuk masing-masing kelompok umur

DATA YANG DIGUNAKAN



```
## Observations: 9,754
## Variables: 5
## $ loan_id      1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1...
## $ tanggal_cair "2020-01-06", "2020-01-06", "2020-01-06", "2020-01-06", ...
## $ cabang       "AA", "AA", "AA", "AA", "AA", "AA", "AB", "AB", "AB", "A...
## $ agen         "AA-1", "AA-1", "AA-1", "AA-2", "AA-2", "AA-2", "AB-1", ...
## $ amount       320000, 440000, 200000, 430000, 360000, 220000, 320000, ...
```

Terlihat bahwa ada 9,754 baris data (Observations) dan ada 5 kolom (Variables),

loan_id : unik ID dari data ini

tanggal_cair : tanggal uang diberikan kepada mitra

cabang : lokasi agen bekerja dan tempat mitra terdaftar

agen : petugas lapangan yang melakukan pencairan

amount : jumlah uang yang dicairkan

1. MEMFILTER DATA YANG DIGUNAKAN

```
### # A tibble: 10 x 2
##   cabang total_amount
##
## 1 AA          75710000
## 2 AB          81440000
## 3 AC          83990000
## 4 AD          76080000
## 5 AE          54200000
## 6 AF          68040000
## 7 AG          74080000
## 8 AH          73840000
## 9 AI          46640000
## 10 AJ          43580000
```

Langkah pertama yang dilakukan adalah *filter* untuk data bulan mei 2020 dan menghitung jumlah uang yang dicairkan (*total amount*) setiap cabang

```
SCRIPT :
library(dplyr)
df_loan_mei <- df_loan %>%
  filter(tanggal_cair >= '2020-05-01', tanggal_cair <= '2020-
05-31') %>%
  group_by(cabang) %>%
  summarise(total_amount = sum(amount))
df_loan_mei
```

2. LIMA CABANG TERBAIK DAN TERBURUK

Langkah kedua yang dilakukan adalah melihat 5 cabang dengan *total amount* paling besar dan paling kecil. Diketahui cabang AC merupakan cabang dengan *total amount* terbesar dan cabang AV merupakan cabang dengan *total amount* terkecil

```
## # A tibble: 5 x 2
##   cabang total_amount
##
## 1 AC      83,990,000
## 2 AB      81,440,000
## 3 AD      76,080,000
## 4 AA      75,710,000
## 5 AG      74,080,000
```

```
## # A tibble: 5 x 2
##   cabang total_amount
##
## 1 AV      30,280,000
## 2 AS      31,740,000
## 3 AT      34,840,000
## 4 AU      35,610,000
## 5 AO      39,120,000
```

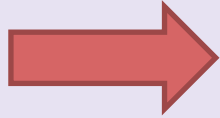
```
SCRIPT :
library(dplyr)
library(scales)

df_loan_mei %>%
  arrange(desc(total_amount)) %>%
  mutate(total_amount =
    comma(total_amount)) %>%
  head(5)

df_loan_mei %>%
  arrange(total_amount) %>%
  mutate(total_amount =
    comma(total_amount)) %>%
  head(5)
```

3a. MENGHITUNG UMUR CABANG (dalam bulan)

```
## # A tibble: 10 x 3
##   cabang pertama_cair umur
##
## 1 AA      2020-01-06      4
## 2 AB      2020-01-06      4
## 3 AC      2020-01-06      4
## 4 AD      2020-01-06      4
## 5 AE      2020-02-03      3
## 6 AF      2020-02-03      3
## 7 AG      2020-02-03      3
## 8 AH      2020-02-03      3
## 9 AI      2020-03-02      2
## 10 AJ     2020-03-02      2
```



```
## # A tibble: 10 x 4
##   cabang pertama_cair umur total_amount
##
## 1 AA      2020-01-06      4      75710000
## 2 AB      2020-01-06      4      81440000
## 3 AC      2020-01-06      4      83990000
## 4 AD      2020-01-06      4      76080000
## 5 AE      2020-02-03      3      54200000
## 6 AF      2020-02-03      3      68040000
## 7 AG      2020-02-03      3      74080000
## 8 AH      2020-02-03      3      73840000
## 9 AI      2020-03-02      2      46640000
## 10 AJ     2020-03-02      2      43580000
```

Langkah ketiga adalah menghitung sudah berapa lama sejak tanggal cair pertama sampai dengan bulan Mei (umur). Kemudian, untuk membandingkan data umur dan performa di bulan mei dilakukan penggabungan data sebelumnya dengan umur.

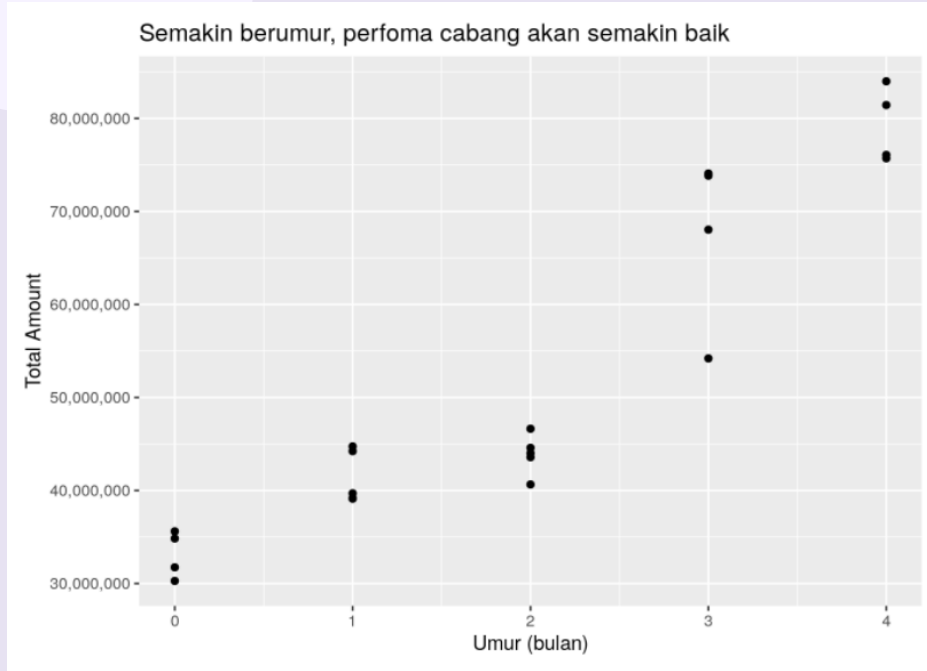
SCRIPT :

```
library(dplyr)
```

```
df_cabang_umur <- df_loan
%>%
  group_by(cabang) %>%
  summarise(pertama_cair =
    min(tanggal_cair)) %>%
  mutate(umur =
    as.numeric(as.Date('2020-05-15')
      - as.Date(pertama_cair)) %/%
    30)
df_cabang_umur
```

```
df_loan_mei_umur <-
df_cabang_umur %>%
  inner_join(df_loan_mei, by =
    'cabang')
df_loan_mei_umur
```

3b. PLOT RELASI UMUR DAN PERFORMA



SCRIPT :
library(ggplot2)

```
ggplot(df_loan_mei_umur, aes(x = umur, y = total_amount)) +  
  geom_point() +  
  scale_y_continuous(labels =  
    scales::comma) +  
  labs(title = "Semakin berumur,  
    performa cabang akan semakin  
    baik",
```

```
x = "Umur (bulan)",  
y = "Total Amount")
```

Pada tahap ini kita membuat **plot antara umur dan performa**. Berdasarkan *output* di atas dapat kita simpulkan bahwa **semakin berumur suatu cabang maka performanya juga semakin tinggi atau semakin baik**.

4a. MENCARI CABANG YANG PERFORMA RENDAH SETIAP UMUR

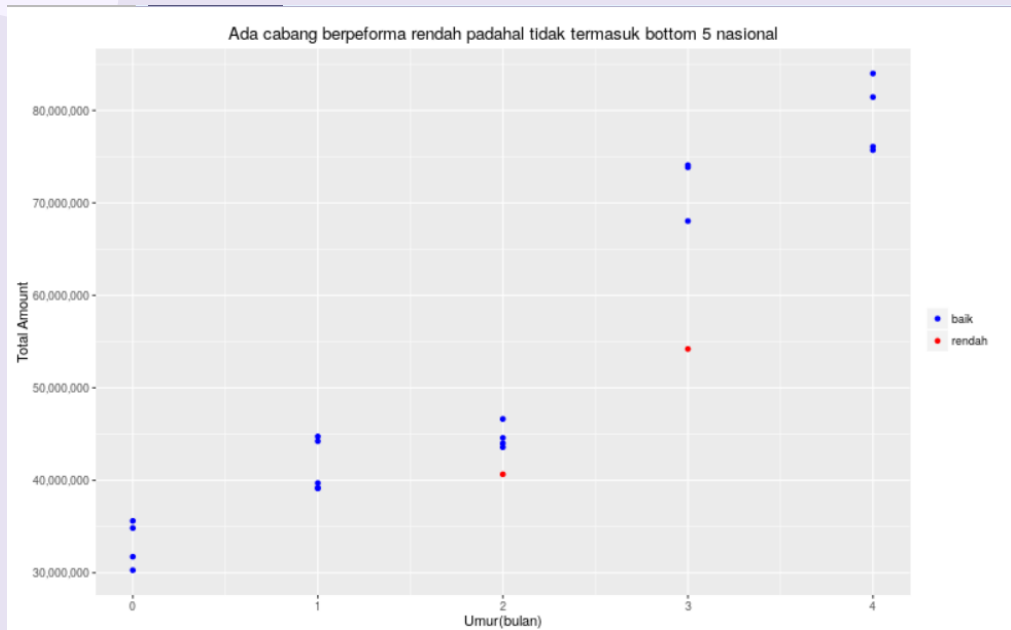
```
## # A tibble: 2 x 8
## # Groups:   umur [2]
##   cabang pertama_cair umur total_amount Q1          Q3          IQR          flag
##
## 1 AE      2020-02-03      3 54,200,000  64,580,000 73,900,000 9,320,000 rendah
## 2 AL      2020-03-02      2 40,650,000  43,580,000 44,590,000 1,010,000 rendah
```

Performa cabang dikatakan **rendah** apabila **performa kurang dari (Q1-IQR)**. Pada *output* di atas, diketahui bahwa terdapat **2 cabang** yang memiliki **performa rendah**. Pada **umur 3 bulan** adalah cabang **AE** dan pada **umur 2 bulan** adalah cabang **AL**.

```
SCRIPT :
library(dplyr)
library(scales)
df_loan_mei_flag <-
df_loan_mei_umur %>%
  group_by(umur) %>%
  mutate(Q1 =
quantile(total_amount, 0.25),
          Q3 =
quantile(total_amount, 0.75),
          IQR =
(Q3-Q1)) %>%
  mutate(flag =
ifelse(total_amount < (Q1 - IQR),
'rendah','baik'))

df_loan_mei_flag %>%
  filter(flag == 'rendah') %>%
  mutate_if(is.numeric,
funs(comma))
```


4b. SCATTER PLOT CABANG YANG MEMILIKI PERFORMA RENDAH



```
SCRIPT :  
library(dplyr)  
library(scales)  
df_loan_mei_flag <-  
df_loan_mei_umur %>%  
  group_by(umur) %>%  
  mutate(Q1 =  
    quantile(total_amount, 0.25),  
           Q3 =  
    quantile(total_amount, 0.75),  
           IQR =  
    (Q3-Q1)) %>%  
  mutate(flag =  
    ifelse(total_amount < (Q1 - IQR),  
    'rendah', 'baik'))  
  
df_loan_mei_flag %>%  
  filter(flag == 'rendah') %>%  
  mutate_if(is.numeric,  
    funs(comma))
```

Pada *output* diatas, kita lebih mudah melihat cabang dengan performa yang rendah dengan tanda berwarna merah.

4c. MENGANALISIS MENGAPA CABANG AE & AL MEMILIKI PERFORMA RENDAH

```
## # A tibble: 4 x 7
## # Groups:   cabang [4]
##   cabang flag   jumlah_hari agen_aktif total_loan_cair avg_amount total_amount
##
## 1 AE      rendah 21           3          175      309,714.3    54,200,000
## 2 AF      baik   21           3          225      302,400      68,040,000
## 3 AH      baik   21           3          241      306,390      73,840,000
## 4 AG      baik   21           3          241      307,386.9    74,080,000
```

Pada *output* diatas, kita ambil contoh untuk **umur 3 bulan** dengan membandingkan performa pada cabang lain. Dilihat **detail performa pada bulan mei** dengan menghitung jumlah hari pencairan dalam 1 bulan, jumlah agen yang aktif, total *loan* yang cair, dan rata-rata *amount* cair per *loan*. Terlihat terjadi **perbedaan yang sangat jauh pada total *loan* yang cair** di cabang AE daripada cabang lainnya

```
SCRIPT :
library(dplyr)
library(scales)
df_loan_mei_flag %>%
  filter(umur == 3) %>%
  inner_join(df_loan, by = 'cabang')
%>%
  filter(tanggal_cair >= '2020-05-01',
         tanggal_cair <= '2020-05-31') %>%
  group_by(cabang, flag) %>%
  summarise(jumlah_hari =
    n_distinct(tanggal_cair),
             agen_aktif = n_distinct(agen),
             total_loan_cair =
    n_distinct(loan_id),
             avg_amount = mean(amount),
             total_amount = sum(amount))
%>%
  arrange(total_amount) %>%
  mutate_if(is.numeric, funs(comma))
```

4d. MEMBANDINGKAN DETAIL PERFORMA CABANG AE DAN CABANG AH

```
## # A tibble: 3 x 6
## # Groups:   cabang [1]
##   cabang agen jumlah_hari total_loan_cair avg_amount total_amount
##
## 1 AE      AE-3  4.0           16           310,625    4,970,000
## 2 AE      AE-2 18.0           73           320,274   23,380,000
## 3 AE      AE-1 21.0           86           300,581   25,850,000
```

```
## # A tibble: 3 x 6
## # Groups:   cabang [1]
##   cabang agen jumlah_hari total_loan_cair avg_amount total_amount
##
## 1 AH      AH-3 19           74.0           303,649    22,470,000
## 2 AH      AH-1 21           81.0           301,358    24,410,000
## 3 AH      AH-2 21           86.0           313,488    26,960,000
```

Pada *output* disamping. Kita akan membandingkan **performa cabang AE dan cabang AH per agen**. Cabang AH memiliki performa yang **jauh lebih baik** daripada cabang AE. Hal ini disebabkan karena pada cabang AE agen 3 melakukan **pencairan hanya 4 hari dalam 1 bulan**, padahal agen lain bisa **aktif 21 hari**. Hal ini membuat **performa cabang AE menjadi rendah** dibanding cabang lain

SCRIPT AE :

```
library(dplyr)
library(scales)
df_loan_mei_flag %>%
  filter(umur == 3, flag == 'rendah') %>%
  inner_join(df_loan, by = 'cabang') %>%
  filter(tanggal_cair >= '2020-05-01',
         tanggal_cair <= '2020-05-31') %>%
  group_by(cabang, agen) %>%
  summarise(jumlah_hari =
    n_distinct(tanggal_cair),
             total_loan_cair = n_distinct(loan_id),
             avg_amount = mean(amount),
             total_amount = sum(amount)) %>%
  arrange(total_amount) %>%
  mutate_if(is.numeric, funs(comma))
```

SCRIPT AH :

```
library(dplyr)
library(scales)
df_loan %>%
  filter(cabang == 'AH') %>%
  filter(tanggal_cair >= '2020-05-01',
         tanggal_cair <= '2020-05-31') %>%
  group_by(cabang, agen) %>%
  summarise(jumlah_hari =
    n_distinct(tanggal_cair),
             total_loan_cair =
    n_distinct(loan_id),
             avg_amount = mean(amount),
             total_amount = sum(amount))
%>% arrange(total_amount) %>%
  mutate_if(is.numeric, funs(comma))
```

KESIMPULAN & SARAN

1. Melakukan perbandingan performa cabang sebaiknya dikelompokkan berdasarkan karakteristik yang sama. Misalnya saja ketika cabang terletak di berbeda wilayah yang memiliki kultur berbeda.
2. Pembahasan lebih lanjut mengenai analisis ini adalah meneliti kenapa ada agen yang hanya aktif beberapa hari saja dalam sebulan.
3. Perusahaan mampu memastikan agen mampu bekerja aktif setiap harinya, sehingga bisa menjaga performa cabang.



THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**



#DQLABPRJ8 DHHOKI

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Bintang Apriansyah Tambunan

Has Completed in
Project Data Analysis for Finance: Performa Cabang

Sep 18, 2020