

# Check Point & Final Project DBA

## Kelompok 28 - DBA G

- Yuni Nurdiyanti (DBA2202990) : Dept Produk
- Yuni Saputri (DBA2310686) : Dept Human Resource
- Aulia Riqzisyafira (DBA2246760) : Dept Marketing
- Elsa Shahnaz Aprilia (DBA2323846) : Dept Sales

## Daftar Isi

Latar Belakang .....	3
Metodologi .....	4
Hasil dan Pembahasan Dept HR.....	5-7
Hasil dan Pembahasan Dept Sales.....	8-9
Hasil dan Pembahasan Dept Marketing .....	10-12
Hasil dan Pembahasan Dept Product.....	13-18
Saran dan Kesimpulan .....	19

## Latar Belakang

Kami dari perusahaan XYG adalah sebuah toko offline store yang menjual berbagai macam kebutuhan, namun semenjak pandemi XYG mulai beralih ke toko online dan berjualan melalui sebuah platform. Saat ini XYG masih kesulitan untuk menentukan customer segmentation dikarenakan kami menjual semua barang dan tidak terfokus dengan satu seksi customer yang baru kami sadari bahwa untuk menjual secara online harus ada minimal satu target customer yang ingin dicapai sehingga memudahkan pencarian toko dimana itu akan berpengaruh kepada keuntungan toko.

Saat ini tim XYG sedang mengevaluasi bersama antara seluruh department (sales, product, hr, dan marketing) untuk sama-sama menentukan customer segmentation dan beberapa action lainnya.

## Metodologi

Pada saat ini semua department yang ada pada toko XYG mulai mencari permasalahan pada masing-masing department dengan KPI yang sudah ditetapkan oleh masing-masing department.

Maka dari itu selain menggunakan KPI, department satu dengan yang lainnya sangatlah berhubungan, hal ini dilakukan untuk mengidentifikasi bersama bagaimana customer segmentation yang tepat dari pelanggan toko XYG.

Masing-Masing department membuat analisa dengan bantuan beberapa tools yaitu:

- Postgre SQL untuk menggabungkan variabel data yang dibutuhkan
- Python via Google Colab untuk melakukan data understanding, cleansing, exploratory sampai pembuatan model
- Dan, terakhir menggunakan tableau untuk memvisualisasikan datanya sesuai dengan business question yang telah diidentifikasi

### 1. Dapertemen Human Resorces

Dapertemen Human Resources merupakan dapertemen yang manajemen SDM. Tentu saja dalam perusahaan diinginkan SDM yang berkualitas dan memiliki produktivitas kerja yang tinggi sehingga dapat memaksimalkan tujuan ataupun misi perusahaan.

Produktivitas pekerja dapat dilihat dari kehadiran atau absensi. Tinggi rendahnya tingkat absensi berpengaruh terhadap produktivitas kerja karyawan, semakin tinggi tingkat absensi akan menurunkan produktivitas kerja begitu juga semakin rendah tingkat absensi akan meningkatkan produktivitas kerja karyawan. Pendapat ini didukung penelitian B. Anderson & D.J. Geldenhuys (2011) dan Fiona Cocker all (2013) menyatakan bahwa adanya hubungan antara ketidakhadiran (tingkat absensi) dengan produktivitas kerja. Tingkat ketidakhadiran merupakan ketidakhadiran pegawai yang berkenaan dengan tugas dan kewajibannya. Pada umumnya perusahaan selalu memperhatikan pegawainya untuk datang dan pulang tepat waktu, sehingga pekerjaan tidak tertunda. Ketidakhadiran seorang pegawai akan berpengaruh terhadap produktivitas kerja, sehingga instansi atau lembaga tidak bisa mencapai tujuan secara optimal.

Berdasarkan dataset yang digunakan adalah employees dan distribution centers. Untuk data pekerja pada employees terlihat bahwa angka absensi selama satu tahun sangat beragam, sebagai HR tentu saja perlu mengevaluasi mengenai angka absen dan menginginkan tingkat absen yang rendah, dalam hal ini akan dilakukan analisis hubungan antara umur dan absensi untuk mendapatkan kriteria pekerja yang baik sehingga produktivitas kerja terjaga.

### 1. Dapertemen Human Resorces

Dalam melakukan analisis data set, hal pertama yang dilakukan adalah mempersiapkan dataset yang digunakan yaitu dataset empolyees dan distribution centers dengan potgreSQL yakni data join dengan fullouterjoin, selain itu dilakukan data cleansing dan preparation dengan menghapus pekerja dibawah umur 18 tahun dan mengganti tipe data untuk kolom age, length service, dan absent hours menjadi integer.

Data hasil join kemudian diproses dengan Pyhton via google colab, dilakukan cleansing data yakni menghilangkan outliers, selain itu dilakukan exploratory sampai pembuatan model. Model yang digunakan adalah prediction dengan linear regresi. Pada analisis ini bertujuan untuk melihat hal apa yang mempengaruhi tingkat absensi sehingga pada model ini variabel dependen adalah absent hours sedangkan untuk variabel independen yang ditinjau adalah umur. Dengan analisis regresi akan dilihat bagaimana pengaruh umur terhadap lama absen. dari hasil regresi  $y=4.20059465x-115.55161296$ . Dengan nilai akurasi sebesar 71%

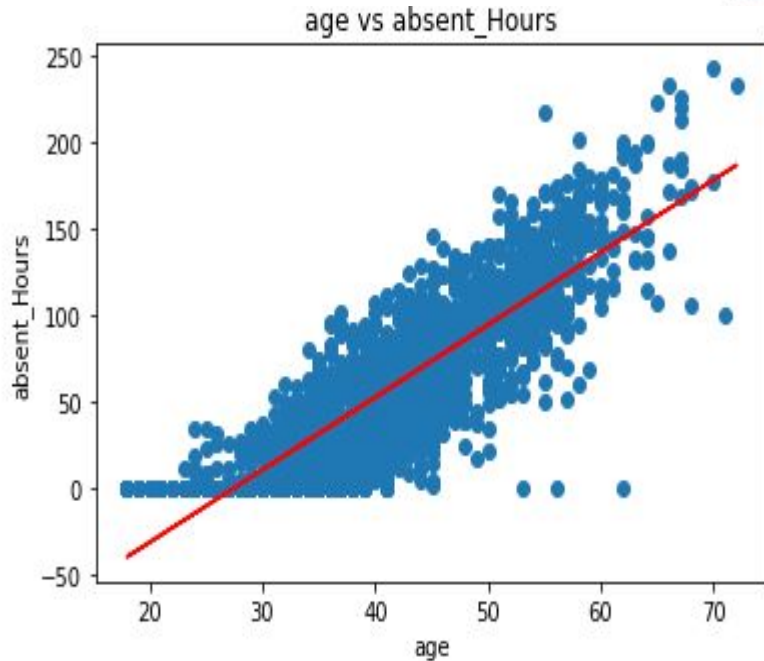
## Hasil dan Pembahasan

### 1. Dapertemen Human Resorces

Dari hasil regresi terlihat bahwa semakin meningkat umur maka, lama absensi juga ikut meningkat. Artinya pekerja dengan umur senja lebih banyak tidak hadir.

Di sisi lain perusahaan menginginkan pekerja produktif dengan nilai absensi rendah, hal ini mungkin dikarenakan pada usia senja stamina cenderung menurun. Oleh karena itu akan lebih baik apabila nantinya perusahaan merekrutmen pekerja dengan usia muda dan menerapkan sop terkait pensiun untuk pekerja

Dapat dilihat pada gambar disamping untuk hasil regresi dengan akurasi 71% dianggap model dapat diterima



## Hasil dan Pembahasan

### 2. Department Sales

Department sales adalah department yang berfokus pada penjualan, pada identifikasi masalah kali ini department sales perlu menemukan pola berbelanja atau customer behaviour dari tahun 2018-2022 sehingga department product mendapatkan insight dari segi persediaan stock barang maupun kualitas barang yang dijual sehingga berpengaruh pada department sales.

Setelah mengidentifikasi kemungkinan masalah dengan membuat business question yang ada pada analysis kali ini, department sales menggunakan data order items dan order yang dijoinkan menggunakan Postgre SQL, lalu setelah itu melakukan cleansing dan exploratory dengan tambahan sedikit visualisasi dengan menggunakan Google Colab. Dan membuat modeling data dengan Clustering modeling bertujuan untuk meminimalisasikan *objective function* yang diset dalam proses *clustering* dengan cara meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya juga bertujuan untuk menemukan grup dalam data, dengan jumlah grup yang diwakili oleh variabel K.

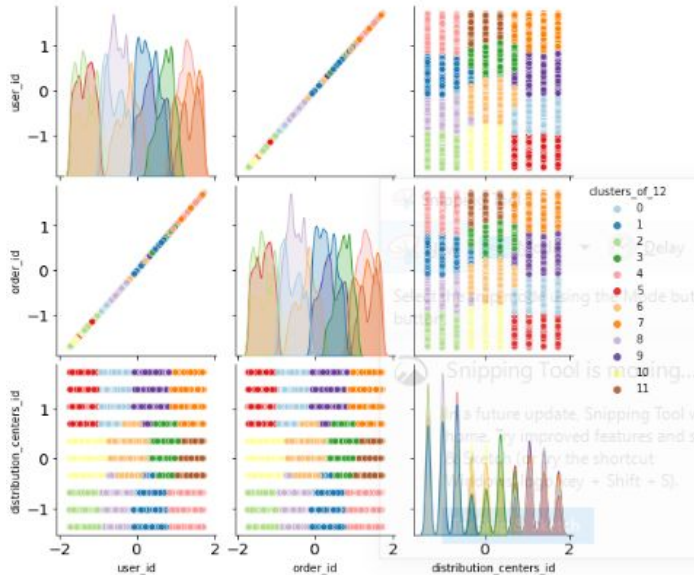
Setelah melakukan clustering guna untuk mengetahui polarisasi dari pembelian customer department sales melakukan visualisasi data dengan Tableau untuk melengkapi jawaban dari business question yang sudah diidentifikasi di awal



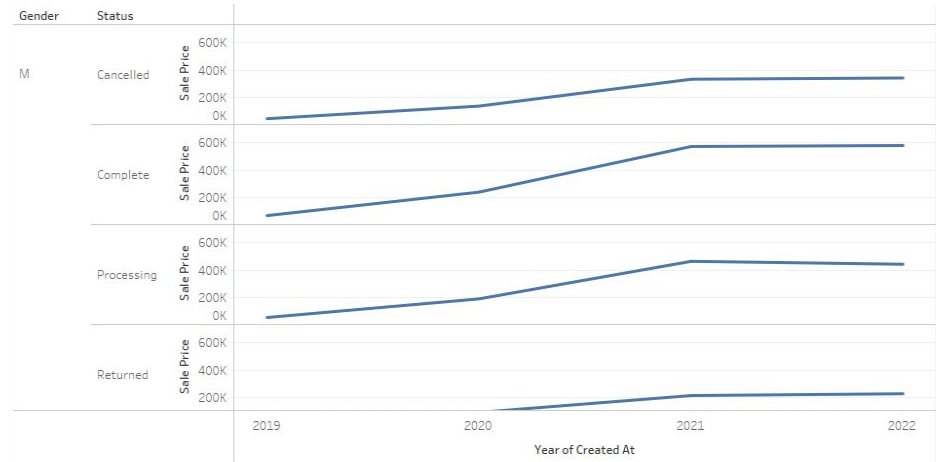
## Hasil dan Pembahasan

### 2. Department Sales

Pada data modeling department sales berhasil menjawab pertanyaan mengenai pola pembelian customer yang dibagi menjadi 5 kelompok begitu juga dengan visualisasi dapat ditemukan bahwa ada beberapa hipotesis dapat dijawab dengan hasil visualisasi, berikut beberapa contoh hasil clustering dan visualisasi data yang dapat dilihat melalui link: <https://public.tableau.com/authoring/ElsaShahnaz-CP5/Statusorderbynumitem#1>



Gender and Status for Create Order by Sale of Price



## Hasil dan Pembahasan

### 3. Department Marketing

Pada Department Marketing ini, kami berfokus pada pemahaman mengenai pola dari pengguna aplikasi sebagai pembeli, pada department ini kami akan memfokuskan pemahaman kami pada 3 tabel yang akan digunakan yaitu tabel users, tabel orders dan juga tabel events. Untuk lebih memahami tentang ketiga tabel tersebut kami melakukan data understanding dan preparation data terlebih dahulu menggunakan python pada google collaboratory, setelah memahami data disini kami menggabungkan table users dan juga table orders dengan menggunakan left join yang ada di SQL dengan tujuan untuk melakukan pengecekan terhadap user / pengguna yang melakukan pembelian. dari pengecekan tersebut kami mendapatkan informasi bahwa dari 100 ribu user terdapat 20.014 user yang tidak atau belum pernah melakukan pemesanan, selain itu disini kami juga mendapatkan bahwa terdapat 5 status pemrosesan barang yaitu Shipped dengan 37466 pesanan, Complete dengan 31269 pesanan, Processing dengan 25110 pesanan, Cancelled dengan 18518 pesanan, Returned dengan 12149 pesanan

## Hasil dan Pembahasan

### 3. Department Marketing

Selain melakukan pengecekan terhadap pengguna yang melakukan pembelian, disini kami juga melakukan pengecekan terhadap keaktifan pengguna dalam menggunakan aplikasi, pengecekan ini melibatkan 2 tabel yaitu tabel users dan juga tabel events yang digabungkan menggunakan left join dan diatur sedemikian rupa sesuai dengan kebutuhan. Gabungan dari kedua tabel ini nantinya digunakan untuk mencari jumlah cluster yang tepat dengan menggunakan metode elbow method dan juga silhouette method, pada pencarian cluster menggunakan kedua metode ini menghasilkan bahwa 3 klaster merupakan jumlah klaster yang tepat untuk dilakukan proses klustering. Setelah ditemukan jumlah cluster yang tepat dilakukanlah pembagian klaster pada data yang nantinya akan dibuat scatter plot dan dilihat tingkat akurasi dari cluster keaktifan pengguna aplikasi dengan cluster pembelian barang.

Hasil dari klustering ini dikatakan bahwa terdapat 84.965 ribu user yang keaktifannya dalam menggunakan aplikasi sama dengan keaktifannya dalam melakukan pembelian barang dan terdapat 15.035 yang berbeda keaktifannya. Maka dari itu dengan adanya penelitian ini diharapkan perusahaan dapat lebih memperhatikan users yang jarang menggunakan aplikasi dan jarang melakukan pembelian barang dengan cara memberikan iklan atau promosi promosi lainnya yang sesuai dengan keinginan mereka

## Hasil dan Pembahasan

### 3. Department Marketing

berikut adalah hasil klastering yang didapat :

	id	nama	cart	purchase	departement	product	home	Cluster
0	1	Christopher Miller	22.0	8.0	22.0	22.0	2.0	1
1	2	Manuel McDonald	0.0	0.0	0.0	0.0	0.0	0
2	3	Michael Reese	2.0	2.0	2.0	2.0	2.0	2
3	4	Devon Gilmore	4.0	2.0	4.0	4.0	0.0	0
4	5	Brandon Hernandez	16.0	4.0	16.0	16.0	0.0	1

berikut adalah hasil akurasi dan juga scatterplot dari klastering yang didapat :



## 4. Departement Product

Department produk merupakan salah satu department dibawah naungan perusahaan PT.XYZ. departemen ini yang mengelola terkait produk dan inventory pada perusahaan.

Pada department ini kami memfokuskan pada dataset inventory, dataset produk, dan dataset distribusi center. Pada dataset ini kami mengolah beberapa variabel yang akan digunakan untuk menyelesaikan suatu permasalahan pada perusahaan. Variabel yang digunakan yaitu cost product, cost inventory, sold at, produk category, price product, price retail product, sebelum mencari informasi. Data yang diperoleh dilakukan pengolahan terlebih dahulu menggunakan berbagai tools seperti postgresql, python dan tableau

#### 4. Departement Product

<b>model Clustering</b>	<b>silhouette score</b>
K-Means	0,5233
<b>model klasifikasi</b>	<b>akurasi</b>
Naive Bayes	99,51%
Decision Tree Classifier	100%
Support Vector machine	100%
K-NN	99,96%

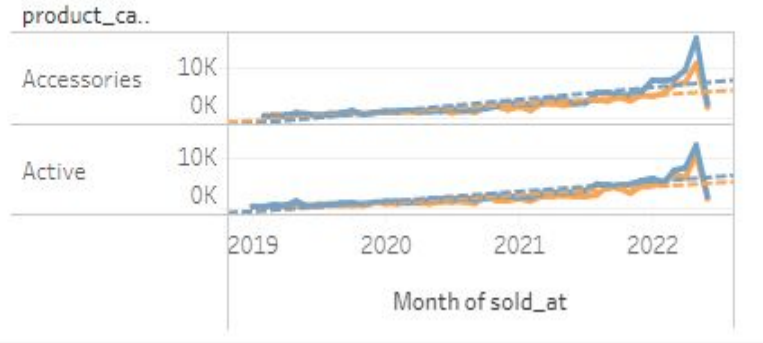
untuk menghasilkan informasi. data i dilakukan pengelompokkan clustering dengan K-Means untuk menentukan cluster persebarannya. hasil cluster yang diperoleh akan digunakan sebagai label target / prediktor data . dengan label target inilah. data akan diklasifikasikan sesuai dengan targetnya menggunakan 4 metode yang berbeda seperti Naive bayes, DTC, SVM dan KNN. hasil klasifikasi menentukan bahwa model DTC dan SVM menghasilkan akurasi yang tinggi sebesar 100%. sehingga untuk model yang dibangun ini cocok digunakan mengolah data pada dept.produk

## hasil dan pembahasan

### prediksi profit



### prediksi penjualan produk sesuai department

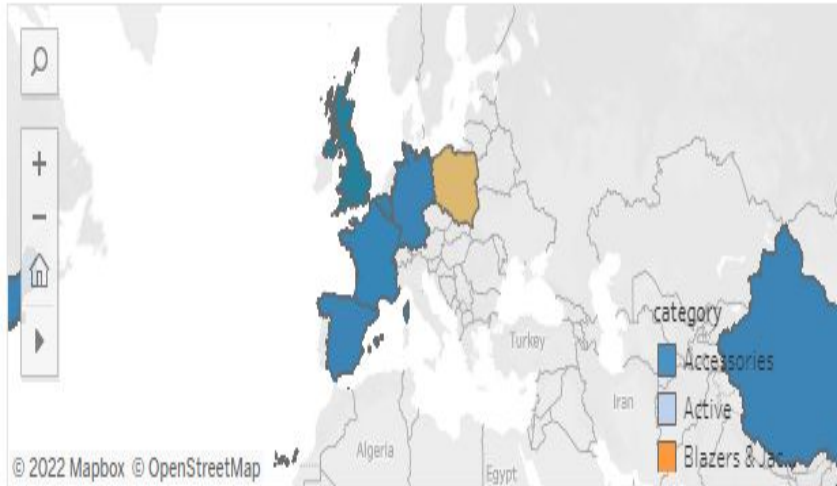


- pada visualisasi prediksi profit ini digunakan untuk melihat prediksi keuntungan pada tahun berikutnya berdasarkan data yang dimiliki. nilai estimasi ini diperoleh dari persebaran trend forecasting pada data aktual pada linear regressi
- pada visualisasi penjualan produk sesuai department ini. melihat banyaknya tren pembelian pada kategori produk tertentu. tren ini dilihat dari 2 sisi yaitu sisi men dan sisi women. setiap kategori produk ini akan dilakukan sorting untuk melihat keuntungan dari hasil tren penjualan.

## Hasil dan Pembahasan

### 4. Departement Product

persebaran negara pembuatan produk



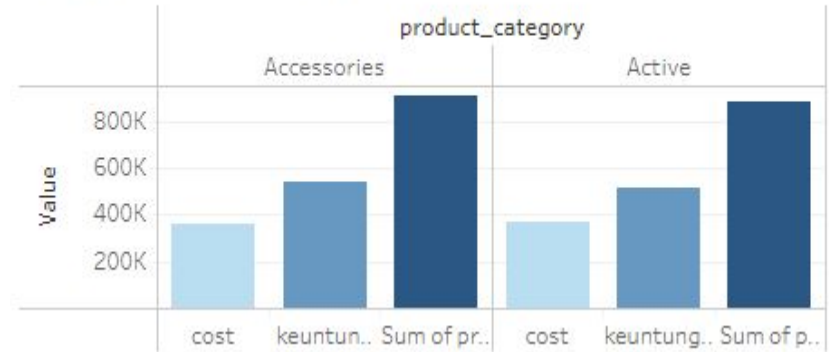
pada gambar disamping kiri merupakan sebuah visualisasi terkait persebaran negara pembuatan category produk. warna yang ditampilkan merupakan hasil filtering kategori yang digunakan, semakin banyak field yang digunakan maka semakin banyak tampilan visualisasi yang diberikan



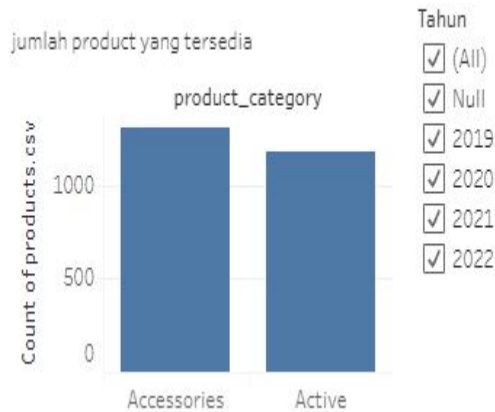
## Saran dan Kesimpulan

pada gambar disamping kanan merupakan sebuah visualisasi terkait diagram keuntungan pada sebuah kategori produk. hasil keuntungan ini diperoleh dari kalkulasi nilai  $\text{sum of price product} - \text{cost}$ . semakin besar nilai  $\text{sum of price product}$  dan  $\text{cost}$  semakin kecil. maka keuntungan yang diperoleh perusahaan akan semakin besar

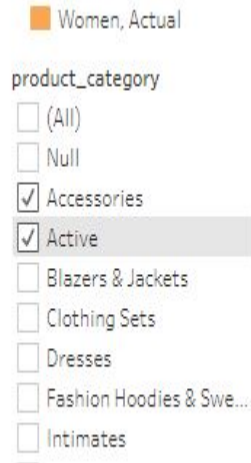
diagram keuntungan



## Saran dan Kesimpulan



pada



pada gambar disamping kiri merupakan sebuah visualisasi terkait jumlah produk yang tersedia. yang mana visualisasi ini sudah menggunakan metode filter data. jika filter data yang dipakai semakin banyak maka jumlah produk yang divisualisasikan semakin banyak

## Saran dan Kesimpulan

Dari setiap metodologi yang dilakukan oleh masing-masing department Toko XYG menemukan bahwa :

- Department HR menemukan bahwa semakin meningkat umur maka, lama absensi juga ikut meningkat. Artinya pekerja dengan umur senja lebih banyak tidak hadir. Hal ini menjadi pertimbangan oleh tim HR untuk memilih para pekerja yang berada pada usia produktif sehingga lebih efektif dalam meningkatkan kualitas perusahaan karena human resource yang baik dalam bekerja
- Department product menemukan bahwa hasil penjualan pada beberapa produk dipengaruhi oleh trend pada tahun tertentu. setiap produk di dominasi oleh laki-laki atau perempuan. sehingga perusahaan dapat menjual produk sesuai dengan keinginan masing masing gender. hasil keuntungan yang diperoleh dari banyaknya pembelian pada suatu produk tertentu yang dikeluarkan oleh perusahaan. dan perusahaan dapat mengeluarkan produk sesuai dengan trend dengan hasil prediksi dari tahun sebelumnya
- Department sales menemukan bahwa harga dan kualitas barang menjadi salah satu faktor mengapa banyak customer yang akhirnya cancel atau melakukan returned, namun dari segi pola pembelian customer XYG dapat terbilang loyal terutama pada quartal 2 dan 4 di hampir setiap bulan, hal ini menjadi informasi yang baik dari department sales untuk lebih memperhatikan pola pembelian customer yang mungkin saja bisa berubah namun dapat menjadi saran untuk department product supaya lebih memperhatikan product yang dijual
- Department Marketing menemukan bahwa terdapat 85% pola penggunaan aplikasi sama dengan pola pembelian barang oleh pelanggan. Maka dari itu diharapkan perusahaan dapat lebih memperhatikan pelanggan yang jarang menggunakan aplikasi dan jarang melakukan pembelian melalui iklan atau promosi-promosi lainnya yang sesuai dengan keinginan mereka.