

Final Project E-Commerce: Mengidentifikasi Profil Pelanggan

Defri Ari Ashadi



Latar Belakang: Di era digitalisasi sekarang banyak sekali masyarakat pengguna gawai mulai dari handphone, tablet, maupun laptop/komputer sebagai daily driver mereka untuk mencari informasi yang luas atau juga sebagai mata pencaharian untuk sehari-hari mereka. Karena sekarang jual-beli barang sudah bisa mereka akses dengan mudah, cepat, dan terpercaya melalui media platform yang begitu banyak hanya di gengaman mereka. Salah satunya E-Commerce pada suatu Marketplace yang sudah banyak di gunakan masyarakat.

Pengidentifikasian profil pelanggan memuat data riwayat transaksi dari **Everything Plus**, yakni sebuah toko daring yang menjual perlengkapan rumah tangga.

Tujuan: Mencari tahu profil pengguna berdasarkan pola dan perilakunya untuk setiap transaksi yang dilakukan pengguna berdasarkan riwayat pembelian.

RFM

RFM

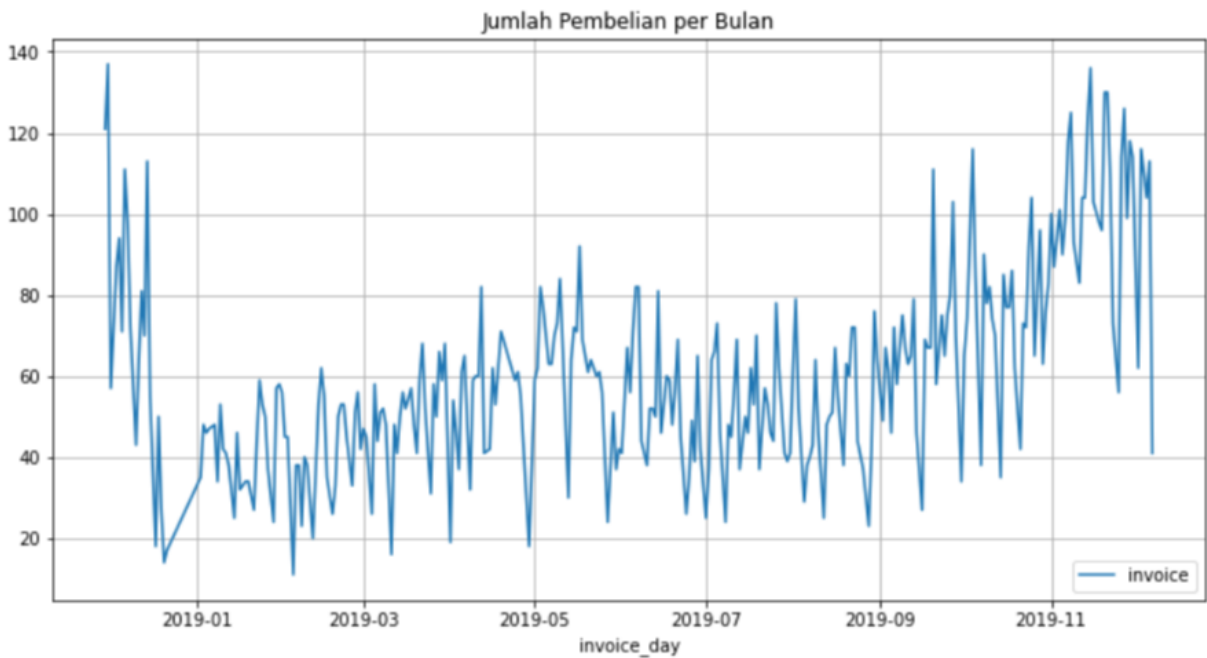
Analisis RFM adalah teknik pemasaran yang digunakan untuk menentukan peringkat dan mengelompokkan pelanggan secara kuantitatif berdasarkan keterkinian, frekuensi, dan total uang dari transaksi terkini mereka untuk mengidentifikasi pelanggan terbaik dan melakukan kampanye pemasaran yang ditargetkan (Katie Terrell Hanna, 2024).

- Recency - Perbedaan antara tanggal referensi dan tanggal maksimum dalam kerangka data untuk setiap pelanggan (yaitu kunjungan terakhir)
- Frequency - Mengetahui seberapa sering pelanggan melakukan pembelian, yang dapat membantu untuk mengidentifikasi pelanggan tetap.
- Monetary - Menunjukkan jumlah total yang dihabiskan oleh pelanggan untuk membeli produk dalam periode waktu tertentu.

Berdasarkan Nilai RFM, skor untuk setiap pelanggan yaitu antara 1 dan 3 (mengelompokkannya). 3 adalah skor terbaik dan 1 adalah skor terburuk (Surendra Tanniru, 2023):

Segment	RFM	Deskripsi
Best Customers	333	Great Recency, Great Frequency, Great Monetary
Loyal Customers	1X1	Good Recency, Frequency, Good Monetary
Big Spenders	XX3	Recency, Frequency, Great Monetary
Casual Customers	XXX	Recency, Frequency, Monetary
Almost Lost	233	Fine Recency, Great Frequency, Great Monetary
Lost Customers	133	Good Recency, Great Frequency, Great Monetary
Lost Cheap Customers	111	Good Recency, Good Frequency, Good Monetary

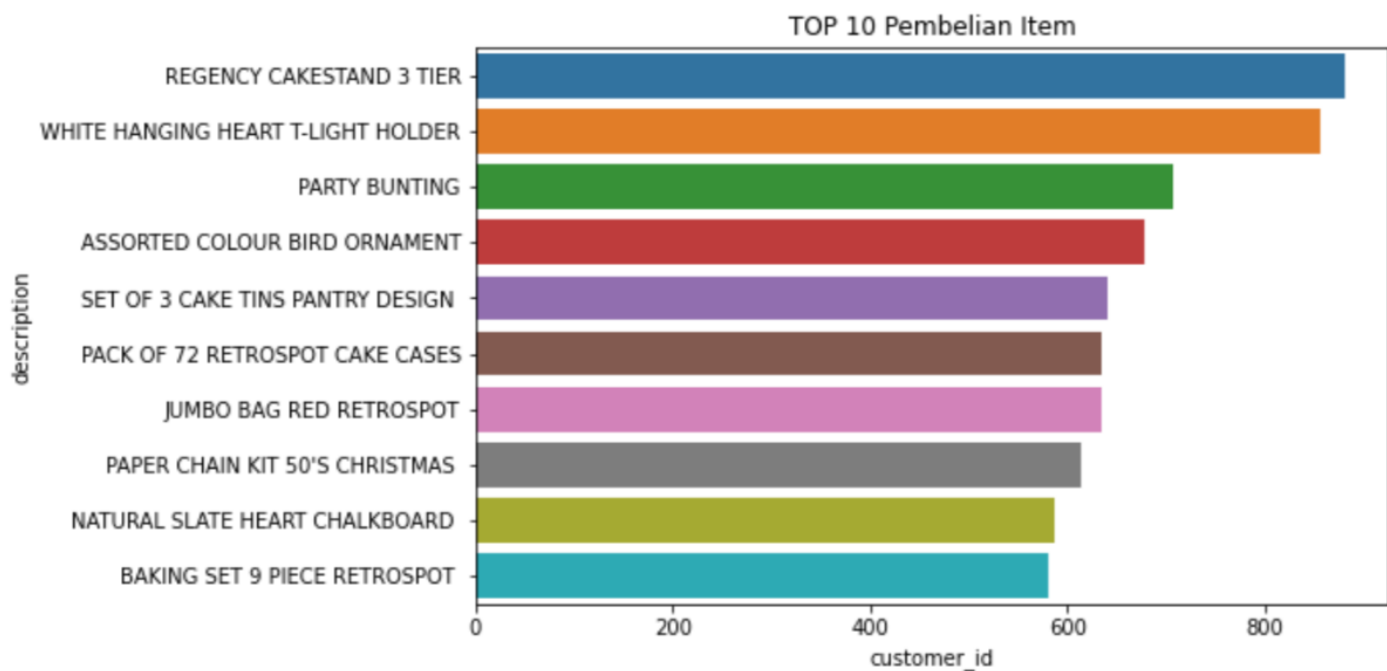
Exploratory Data Analyst



Total transaksi selama satu tahun dalam hari:

- Pada bulan ke-10 terlihat lebih banyak yang melakukan transaksi dengan nilai 175 lebih pelanggan berbeda.

- Pada awal tahun 2018 - 2019 bulan ke-1 terdapat lompatan yang cukup jauh, apakah di waktu tersebut tidak terjadinya transaksi oleh pelanggan?



10 Item yang terjual paling banyak:

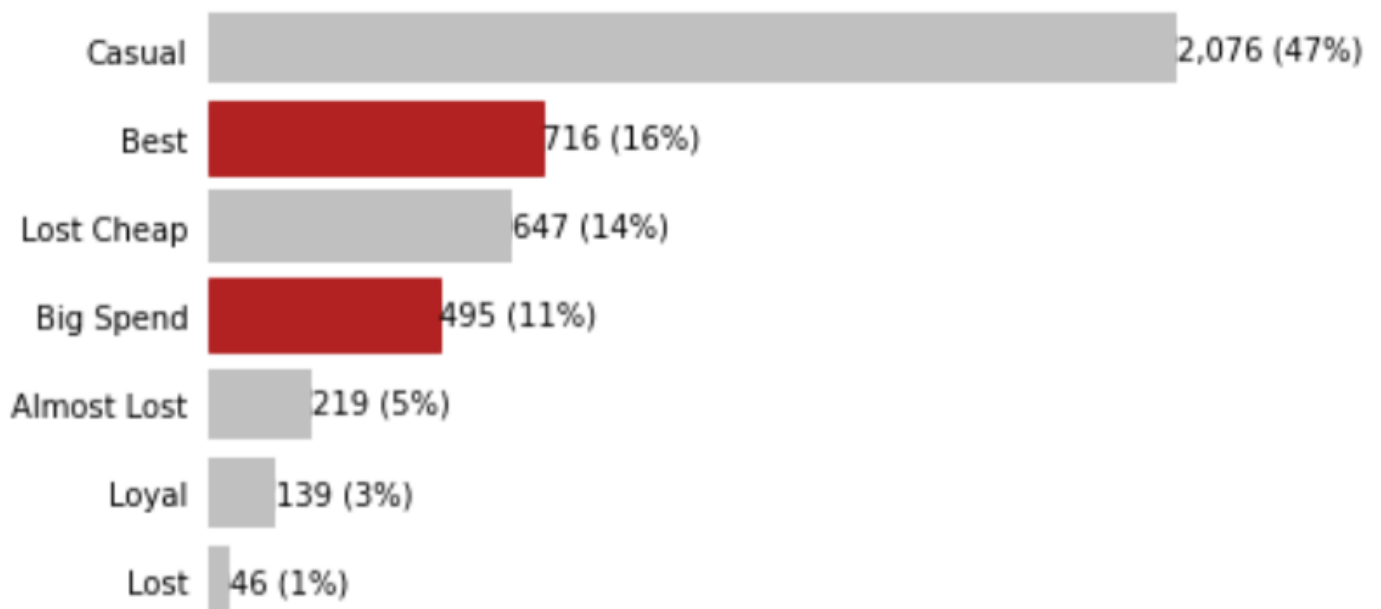
- REGENCY CAKESTAND 3 TIER dan WHITE HANGING HEART T-LIGHT HOLDER merupakan item paling laris dari 800 pelanggan lebih yang membelinya.
- Pada urutan ke-10 ada BAKING SET 9 PIECE RETROSPOT dengan 500 pelanggan lebih.

Segmentasi Pengguna

Segmentasi dilakukan menggunakan nilai Recency, Frequency, dan Monetary dari pelanggan. Untuk membantu kita memperbaiki tepi bin, mari gunakan nilai kuantil kolom "Recency", "Frequency", dan "Monetary Value" (Bala Priya C, 2023).

Quantil(%)	Recency	Frequency	Monetary
30	30	1	384,10
60	90	4	1154,42
90	260	9	3640,84

Customer ID	Recency	Frequency	Monetary	RFM	Segment
12346	1 (326)	1 (1)	3 (77183,6)	113	Big Spend
12347	3 (2)	3 (7)	3 (4310,0)	333	Best
12348	2 (75)	2 (4)	3 (1797,0)	223	Big Spend
12349	3 (19)	1 (1)	3 (1757,0)	313	Big Spend



Segmentasi RFM:

- Terlihat pada segmentasinya ternyata pelanggan dengan segmentasi Casual memiliki jumlah yang paling banyak dengan presentasi di 47% secara keseluruhan.
- Distribusi pelanggan seperti ini cukup normal dan cukup baik, mengingat tingkat lost/almost lost/lost cheap pelanggan lebih sedikit dengan tingkat pelanggan best/big spend/total maupun casual yang masih tinggi.

Uji Hipotesis

Pengujian hipotesis kali ini kita akan menggunakan uji-z, dimana uji ini lebih praktis dibandingkan uji-t yang memiliki nilai kritis berbeda untuk setiap ukuran sampel, sedangkan uji z memiliki nilai kritis tunggal untuk setiap tingkat signifikansi (misalnya, 1,96 untuk 5% dua sisi). Oleh karena itu, untuk ukuran sampel yang besar dan varians populasi yang diketahui, banyak uji statistik dilakukan menggunakan uji z hipotesis (Angel Das, 2022). Pada pengujian hipotesis statistik kali ini merujuk pada:

- Setiap segmentasi pengguna (disini kita pilih 3 segment Best, Big Spend, dan Casual).
- 4 Jenis Item dari ketiga segment yang paling populer yaitu "WHITE HANGING HEART T-LIGHT HOLDER", "REGENCY CAKESTAND 3 TIER", "PARTY BUNTING", dan "ASSORTED COLOUR BIRD ORNAMENT".
- Invoice dari customers

	Big Spend	Casual
	<ul style="list-style-type: none">• H_0 = Apabila rata-rata pembelian item tertentu pada segment Best dan Big Spend Sama• H_1 = Apabila rata-rata pembelian item tertentu pada segment Best dan Big Spend Tidak Sama	<ul style="list-style-type: none">• H_0 = Apabila rata-rata pembelian item tertentu pada segment Best dan Casual Sama• H_1 = Apabila rata-rata pembelian item tertentu pada segment Best dan Casual Tidak Sama
Best	<ul style="list-style-type: none">• Jenis Item: WHITE HANGING HEART T-LIGHT HOLDER, Segment: Best & Big Spend <p>p-value: 0.046365710930915016</p> <ul style="list-style-type: none">• Jenis Item: REGENCY CAKESTAND 3 TIER, Segment: Best & Big Spend <p>p-value: 2.0296704306144164e-06</p> <p>Menolak hipotesis nol: Ada perbedaan yang signifikan di antara proporsinya</p>	<ul style="list-style-type: none">• Jenis Item: PARTY BUNTING, Segment: Best & Casual <p>p-value: 0.03849648629764424</p> <p>Menolak hipotesis nol: Ada perbedaan yang signifikan di antara proporsinya</p>

Poin penting uji Hipotesis:

- Jenis item holder dan bunting hipotesisnya adalah Gagal untuk menolak hipotesis nol baik untuk kelompok Casual dengan Big Spend terhadap segment Best customers.
- Sedangkan pada jenis item cakestand dan ornament yang dimana kita Menolak hipotesis nol.
- Pada kelompok Best, pelanggan lebih banyak melakukan transaksi daripada kelompok Casual untuk jenis item cakestand dan ornament. Berlaku sebaliknya bagi kelompok pada jenis item holder dan bunting.

Kesimpulan Umum

Date: Apr 15, 2024

Dataset df memiliki total 541909 baris data dan 7 kolom. Kita eliminasi data duplikat beserta nilai missing yang meskipun terdapat 25% nilai missing didalamnya akan tetapi data tersebut nantinya juga tidak dapat digunakan untuk menganalisis profil pelanggan.

Berdasarkan hasil dari beberapa poin temuan dan hasil hipotesis kita bisa pastikan untuk target pemasaran yang tepat sasaran yaitu pada segmentasi Best Customers yang dimana cenderung melakukan pembelian untuk item seperti WHITE HANGING HEART T-LIGHT HOLDER sebagai komoditi unggulan. Akan tetapi berdasarkan jumlah transaksinya dan total biaya pengeluaran pelanggan tertinggi bisa juga di fokuskan pada item REGENCY CAKESTAND 3 TIER yang dimana memiliki 289 pelanggan unik dan 12288 dolar total unit price pada riwayat pembeliannya walaupun kuantitasnya sendiri cukup kecil. Sepertinya kita perlu memberikan apresiasi juga terhadap pelanggan terbaik kita agar tetap terjaga statusnya dan bila perlu kita berikan cashback untuk pembelian berikutnya guna untuk bisa lebih sering lagi untuk melakukan transaksi di toko online kita.

Daftar Pustaka

- Angel Das, 2022, <https://towardsdatascience.com/z-test-statistics-formula-python-implementation-3755d67ba0e7>
- Bala Priya C, 2023, <https://www.kdnuggets.com/customer-segmentation-in-python-a-practical-approach>
- Surendra Tanniru, 2023, <https://www.analyticsvidhya.com/blog/2021/07/customer-segmentation-using-rfm-analysis/>
- Katie Terrell Hanna, 2024, <https://www.techtarget.com/searchdatamanagement/definition/RFM-analysis#:~:text=RFM%20analysis%20is%20a%20marketing,and%20perform%20targeted%20marketing%20campaigns.>
- TripleTen, 2024, <https://tripleten.com/trainer/data-analyst/lesson/617e7139-8923-4b05-9cf9-041ba03d0047/task/464cfc6a-06e1-43d4-8f53-d54f60df6359/?from=program>