

Online Shoppers Purchasing Intention

Dokumen
Laporan Final
Project Stage 1 -
Anaconda



Latar Belakang Masalah

- Pada tahun 2021, Jumlah konsumen yang berbelanja online meningkat mencapai 88% dibandingkan tahun 2020.
- Namun, perusahaan ecommerce Anaconda hanya mendapat revenue dari 15% total pengunjung website.
- Perusahaan Anaconda harus meningkatkan purchase rate yaitu jumlah pengunjung website yang menghasilkan revenue per total pengunjung website agar tidak kalah dari kompetitor lain.

Goals & Objective

- Untuk meningkatkan purchase rate menjadi lebih dari **20%** maka harus menganalisis data perilaku-perilaku pengunjung website untuk menemukan insights sebagai acuan untuk memberikan rekomendasi aksi.
- Serta membuat machine learning model untuk memprediksi apakah seorang pengunjung website akan menghasilkan revenue atau tidak.

Exploratory Data Analysis

1. Descriptive Statistics

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12330 entries, 0 to 12329
Data columns (total 18 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   administrative                        12330 non-null  int64
1   administrative_duration              12330 non-null  float64
2   informational                        12330 non-null  int64
3   informational_duration              12330 non-null  float64
4   product_related                     12330 non-null  int64
5   product_related_duration            12330 non-null  float64
6   bounce_rates                        12330 non-null  float64
7   exit_rates                          12330 non-null  float64
8   page_values                         12330 non-null  float64
9   special_day                         12330 non-null  float64
10  month                               12330 non-null  object
11  operating_systems                  12330 non-null  int64
12  browser                           12330 non-null  int64
13  region                            12330 non-null  int64
14  traffic_type                      12330 non-null  int64
15  visitor_type                      12330 non-null  object
16  weekend                            12330 non-null  bool
17  revenue                          12330 non-null  bool

dtypes: bool(2), float64(7), int64(7), object(2)
memory usage: 1.5+ MB
```

Pengamatan

- Jumlah observasi ada 12330 baris dengan 10 kolom numerik dan 7 kolom kategori dan 1 target (revenue)
- Tidak ada missing values di setiap kolom. Jumlah non-null baris sama dengan jumlah observasi.
- Data type untuk setiap kolom sudah konsisten kecuali operating systems, browser, region, dan traffic type yang berupa integer. Keempat kolom tersebut seharusnya memiliki data type string atau object.

Exploratory Data Analysis

1. Descriptive Statistics

	count	mean	std	min	25%	50%	75%	max	no_unique
administrative	12330.0	2.315	3.322	0.0	0.000	1.000	4.000	27.000	27
administrative_duration	12330.0	80.819	176.779	0.0	0.000	7.500	93.256	3398.750	3335
informational	12330.0	0.504	1.270	0.0	0.000	0.000	0.000	24.000	17
informational_duration	12330.0	34.472	140.749	0.0	0.000	0.000	0.000	2549.375	1258
product_related	12330.0	31.731	44.476	0.0	7.000	18.000	38.000	705.000	311
product_related_duration	12330.0	1194.746	1913.669	0.0	184.138	598.937	1464.157	63973.522	9551
bounce_rates	12330.0	0.022	0.048	0.0	0.000	0.003	0.017	0.200	1872
exit_rates	12330.0	0.043	0.049	0.0	0.014	0.025	0.050	0.200	4777
page_values	12330.0	5.889	18.568	0.0	0.000	0.000	0.000	361.764	2704
special_day	12330.0	0.061	0.199	0.0	0.000	0.000	0.000	1.000	6

	count	unique	top	freq	frac_top
month	12330	10	May	3364	27.283
operating_systems	12330	8	2	6601	53.5361
browser	12330	13	2	7961	64.5661
region	12330	9	1	4780	38.7672
traffic_type	12330	20	2	3913	31.7356
visitor_type	12330	3	Returning_Visitor	10551	85.5718
weekend	12330	2	False	9462	76.7397
revenue	12330	2	False	10422	84.5255
special_day	12330	6	0.0	11079	89.854

- Semua kolom memiliki nilai summary yang cukup wajar walaupun seluruh variabel numerik memiliki nilai mean yang berbeda jauh dengan median karena metrik pengukuran memungkinkan didominasi nilai 0 apabila tidak ada aktivitas visitor.
- Kolom special day hanya memiliki 6 unique values dan lebih baik di-treat sebagai tipe kategorikal.
- Kolom special day dan visitor type memiliki salah satu kategori yang terlalu mendominasi di atas 80% sehingga bisa tidak disertakan pada saat modelling.

Exploratory Data Analysis

Kesimpulan Descriptive Statistics

A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?

- Hampir seluruh kolom memiliki tipe data yang sesuai kecuali operating systems, browser, region, dan traffic type yang seharusnya bertipe string atau object karena keempat kolom tersebut bertipe kategorikal.

B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?

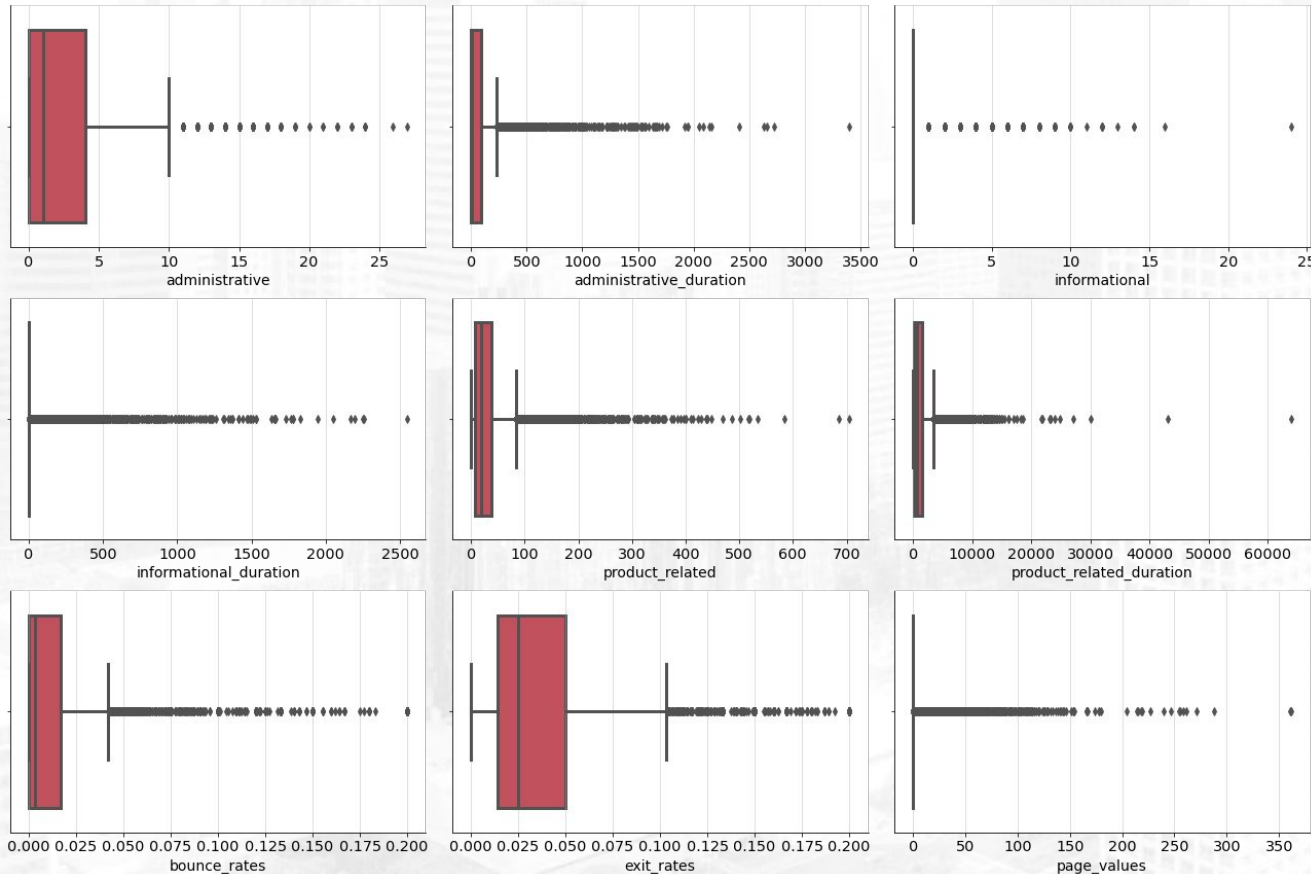
- Semua kolom tidak memiliki missing values.

C. Apakah ada kolom yang memiliki nilai summary agak aneh? (min/mean/median/max/unique/top/freq)

- Semua kolom memiliki nilai summary yang cukup wajar walaupun seluruh variabel numerik memiliki nilai mean yang berbeda jauh dengan median karena metrik pengukuran memungkinkan didominasi nilai 0 apabila tidak ada aktivitas visitor.
- Kolom special day hanya memiliki 6 unique values dan lebih baik di-treat sebagai tipe kategorikal.
- Kolom special day dan visitor type memiliki salah satu kategori yang terlalu mendominasi di atas 80% sehingga bisa tidak disertakan pada saat modelling.
- Selain itu, class imbalance terdapat di kolom target revenue dengan ratio 85:15.

Exploratory Data Analysis

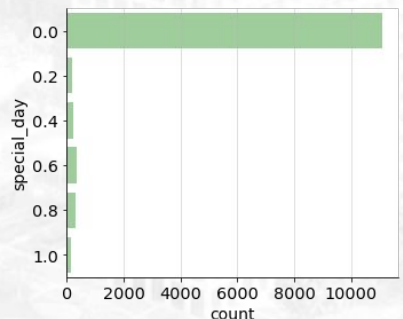
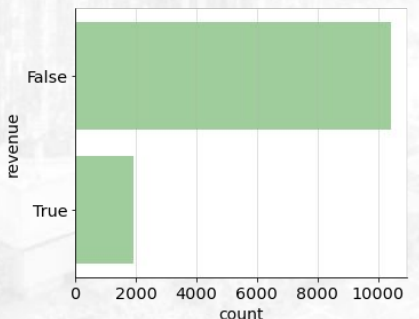
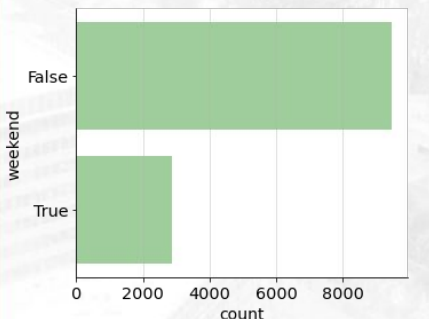
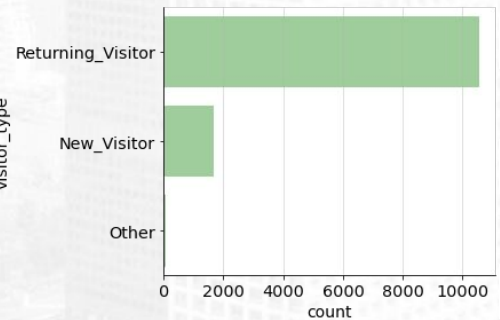
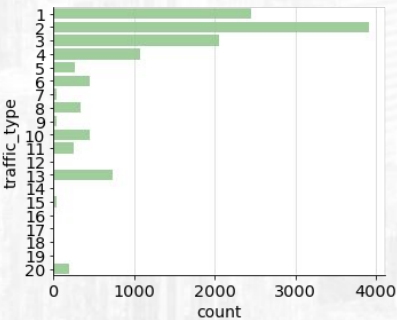
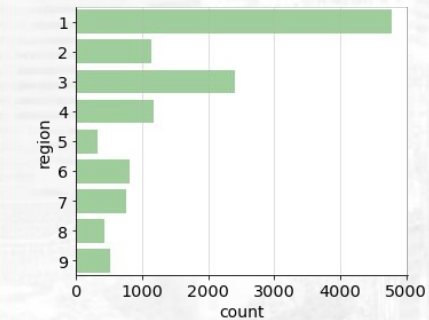
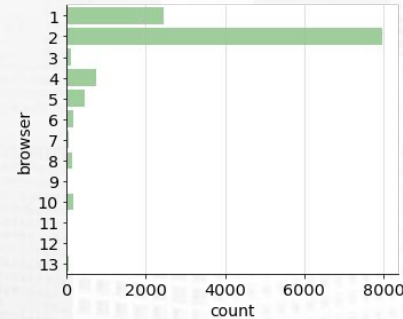
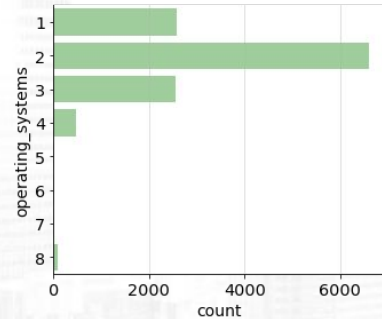
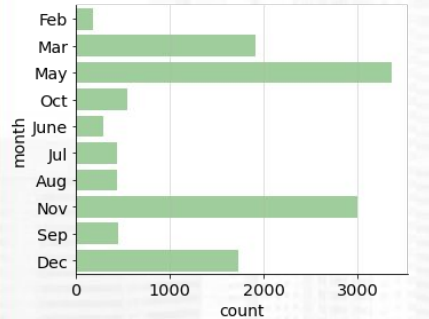
2. Univariate Analysis (Kolom Numerik)



- Berdasarkan plot di atas, seluruh kolom numerik memiliki distribusi yang right-skewed yang mana mayoritas data bernilai 0 dan berada dekat nilai 0.
- Seluruh kolom terlihat memiliki outlier yang cukup banyak
- Pada kolom informational, informational_duration, dan page_values masing-masing quantile terlalu berdekatan. Hal ini sesuai dengan descriptive statistics di atas yang menunjukkan sampai dengan 75% data masih bernilai 0.

Exploratory Data Analysis

2. Univariate Analysis (Kolom Kategorikal)



- Pada kolom bulan, tidak terdapat data pada bulan januari dan april.
- Kolom operating_system dan traffic_type hanya didominasi oleh kategori 1, 2, dan 3, sedangkan kolom browser hanya kategori 1 dan 2.
- Kolom dengan jumlah unique value terbanyak adalah traffic type dan browser.
- Untuk kolom visitor_type ada jenis visitor yang tidak diketahui ditandai oleh Other.
- Berdasarkan target yaitu kolom revenue, terdapat hampir 2,000 pengunjung website menghasilkan revenue (15%) sedangkan lebih dari 10,000 pengunjung website tidak (85%).

Exploratory Data Analysis

Rekomendasi Preprocessing

1. Kolom numerik

- Semua kolom numerik memiliki distribusi right-skewed sehingga butuh ditransformasi agar mendekati distribusi normal.
- Beberapa **metode transformasi** yang bisa digunakan sebagai berikut:
 - log-transformer (`np.log1p`). Alasan memilih `np.log1p` dibandingkan `np.log` karena kolom numerik didominasi nilai 0. Sehingga, apabila menggunakan `np.log` maka hasil transformasi akan bernilai infinity.
 - yeo-johnson method
- Seluruh kolom numerik tidak berada di rentang yang sama sehingga harus diskala ulang atau dinormalisasi.
- Beberapa metode **normalisasi atau scaling** dari scikit-learn yang bisa digunakan yaitu
 - `MinMaxScaler`
 - `StandardScaler`
 - `RobustScaler`
- Outlier handling tidak diterapkan karena secara natural distribusi semua kolom numerik memang skewed.

Exploratory Data Analysis

Rekomendasi Preprocessing berdasarkan Univariate Analysis

2. Kolom kategorikal

- **month**: sin-cos transformation. Dengan transformer ini, hubungan siklus dapat diketahui oleh algoritma machine learning.
- **operating_systems**: ada beberapa opsi data preprocessing yaitu
 - rare label dan one-hot encoder. Kategori 1, 2, dan 3 akan tetap dipertahankan dan selain ketiga kategori itu akan dijadikan satu kategori yaitu other. Kemudian keempat kategori tersebut di one-hot encoded.
 - target encoder atau mean encoder. Seluruh unique value dari kolom ini diubah menjadi fraksi jumlah revenue per jumlah seluruh sesi.
- **browser, region, traffic_type**:
 - rare label dan one-hot encoder
 - target encoder atau mean encoder
- **visitor_type, weekend**: karena unique value hanya 2-3 saja, maka one-hot encoder bisa diterapkan.
- **revenue**: karena terdapat class imbalance dengan ratio 85:15 (majority class: minority class) maka ada beberapa cara preprocessing yang dapat digunakan sebagai berikut
 - resampling (oversampling, undersampling, atau over-under sampling)
 - memberi class weight yang lebih tinggi ke minority class

Exploratory Data Analysis

3. Multivariate Analysis (Analisis Korelasi)

Analisis Korelasi

Analisis korelasi memiliki tujuan untuk mengetahui korelasi antar feature serta korelasi antara feature terhadap label. Berikut ini cara untuk menemukan korelasi setiap pasangan tipe variabel yang akan diuji:

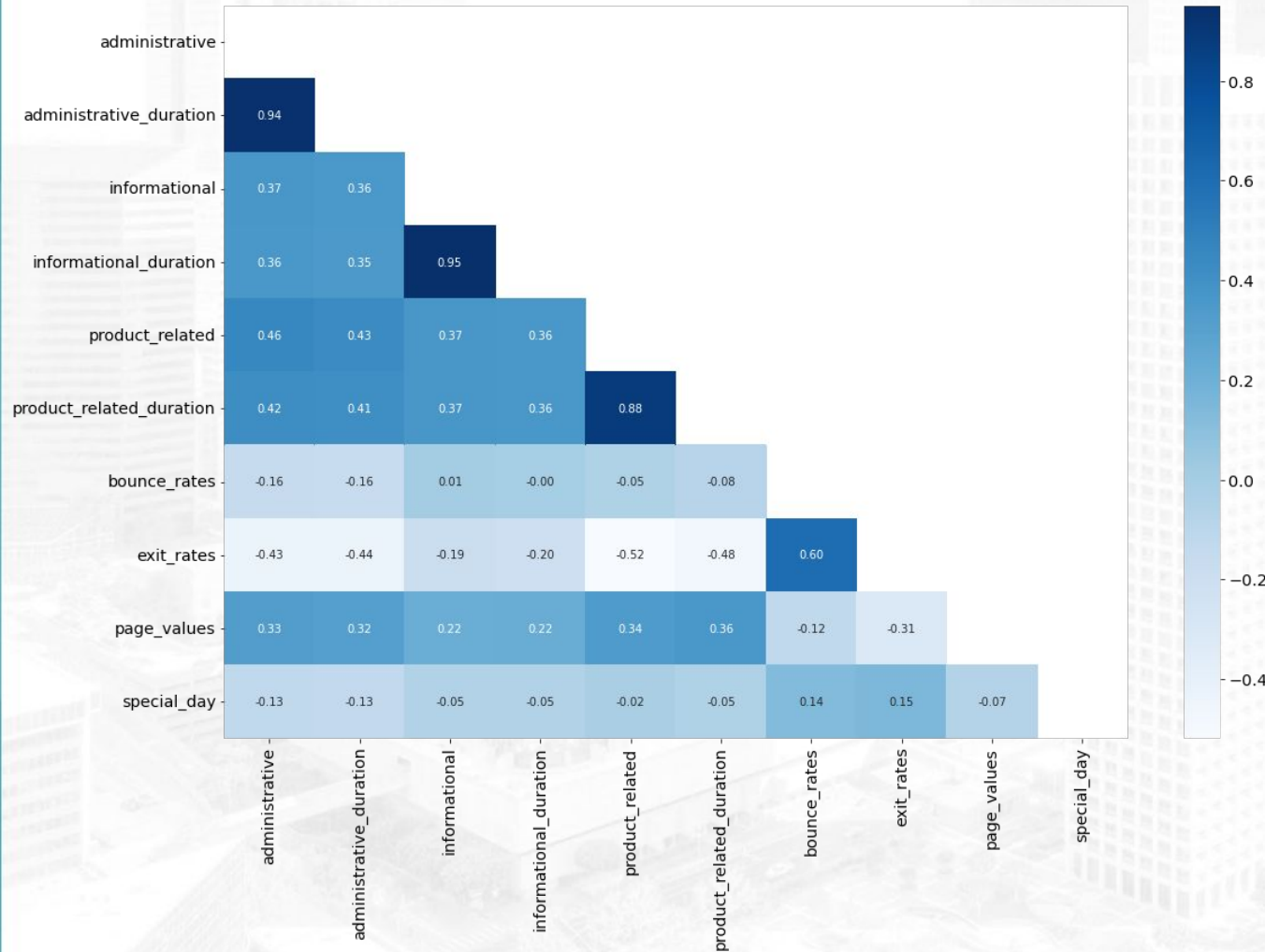
1. **Feature Numerik - Feature Numerik:** spearman's correlation
2. **Feature Numerik - Label Kategorikal:** mann-whitney u test
3. **Feature Kategorikal - Label Kategorikal:** chi-square test

Referensi

- [Exploratory Data Analysis for Feature Selection, Google Cloud](#)
- [Spearman's rho](#)
- [Mann-whitney U test](#)
- [Chi-square test](#)

Exploratory Data Analysis

3. Multivariate Analysis (Korelasi Antar Feature Numerik)



- Kolom jumlah page dan jumlah durasi page memiliki korelasi yang tinggi yaitu 0.88 - 0.95.
- Kolom-kolom numerik yang berkorelasi sangat tinggi harus ditinggalkan salah satunya.
- Selain keempat kolom tersebut, korelasi antar feature numerik dibawah 0.8.

Exploratory Data Analysis

3. Multivariate Analysis (Korelasi Feature Numerik dan Label)

	categorical	group_1	group_2	numerical	statistic	p_value
8	revenue	False	True	page_values	2718419.0	0.000000e+00
7	revenue	False	True	exit_rates	13981307.0	1.156824e-175
5	revenue	False	True	product_related_duration	6502463.0	5.332268e-128
4	revenue	False	True	product_related	6792127.0	1.052439e-107
0	revenue	False	True	administrative	7421135.5	1.311932e-77
1	revenue	False	True	administrative_duration	7487115.0	4.672261e-74
6	revenue	False	True	bounce_rates	12198493.0	1.832770e-61
2	revenue	False	True	informational	8648742.0	1.202781e-36
3	revenue	False	True	informational_duration	8711649.0	1.409039e-35
9	revenue	False	True	special_day	10665043.0	5.079560e-22

- Dari hasil U-test di atas, seluruh kolom memiliki p-value yang lebih kecil dari alpha ($p\text{-value} < \alpha$) sehingga semua kolom numerik memiliki korelasi terhadap label revenue.
- page values, exit rates, product related duration, dan product related memiliki korelasi paling tinggi di antara seluruh kolom numerik.
- Karena p-value product related duration dan exit rates lebih rendah dari product related dan bounce rates, maka kedua variabel pertama yang akan diambil untuk modelling.

Exploratory Data Analysis

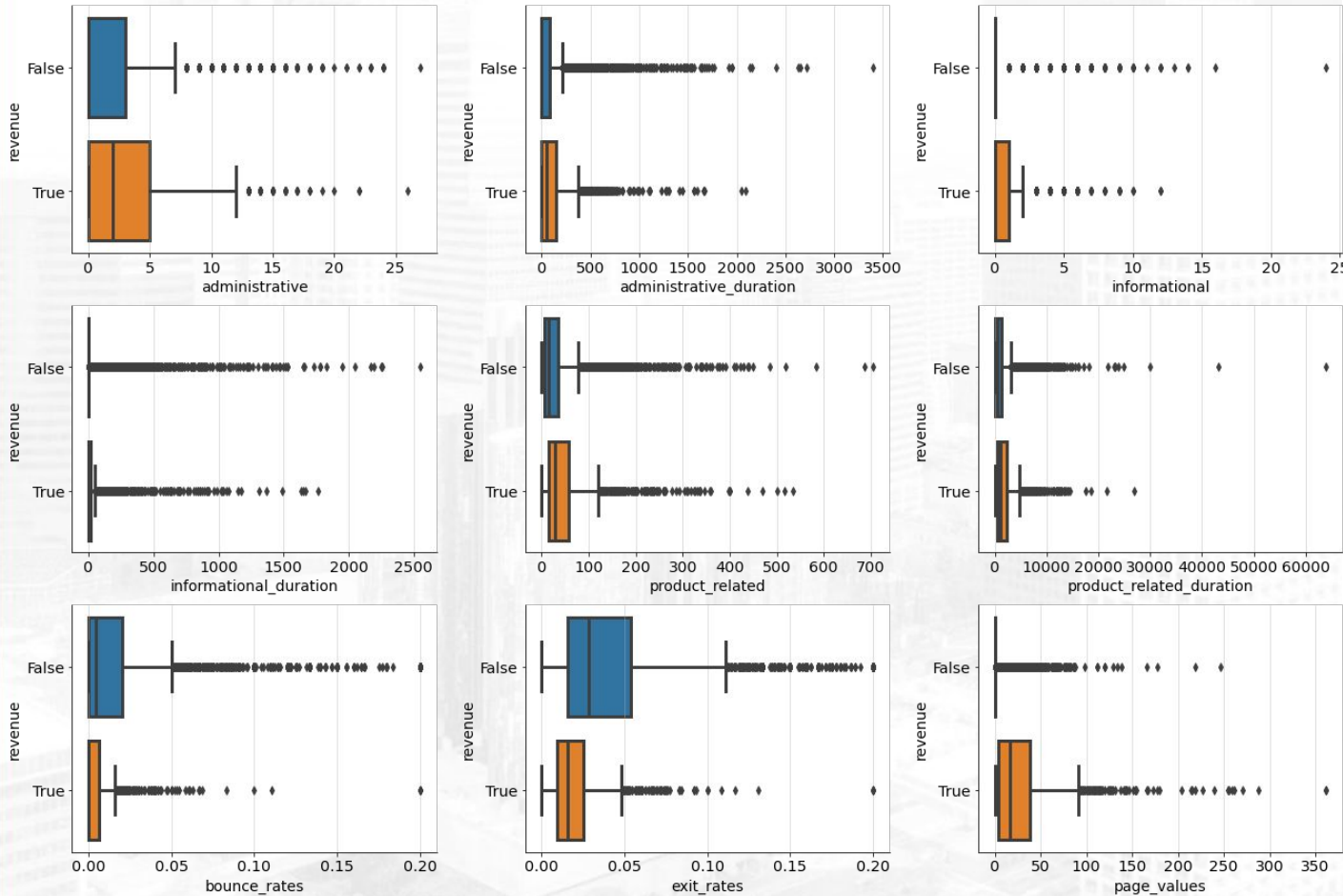
3. Multivariate Analysis (Korelasi Feature Kategorikal dan Label)

	cat1	cat2	chi_stats	p_value	dof	association
0	month	revenue	384.934762	2.238786e-77	9	dependent
4	traffic_type	revenue	373.145565	1.652735e-67	19	dependent
5	visitor_type	revenue	135.251923	4.269904e-30	2	dependent
7	special_day	revenue	96.076906	3.543244e-19	5	dependent
1	operating_systems	revenue	75.027056	1.416094e-13	7	dependent
6	weekend	revenue	10.390978	1.266325e-03	1	dependent
2	browser	revenue	27.715299	6.087543e-03	12	dependent
3	region	revenue	9.252751	3.214250e-01	8	independent

- Berdasarkan hasil chi-square test, region tidak memiliki korelasi terhadap revenue sehingga bisa di-drop saat modelling karena hanya region yang memiliki p-value diatas alpha 0.05.
- Apabila melihat p-value maka kolom kategorikal yang memiliki korelasi paling tinggi adalah month, traffic_type, dan visitor_type

Exploratory Data Analysis

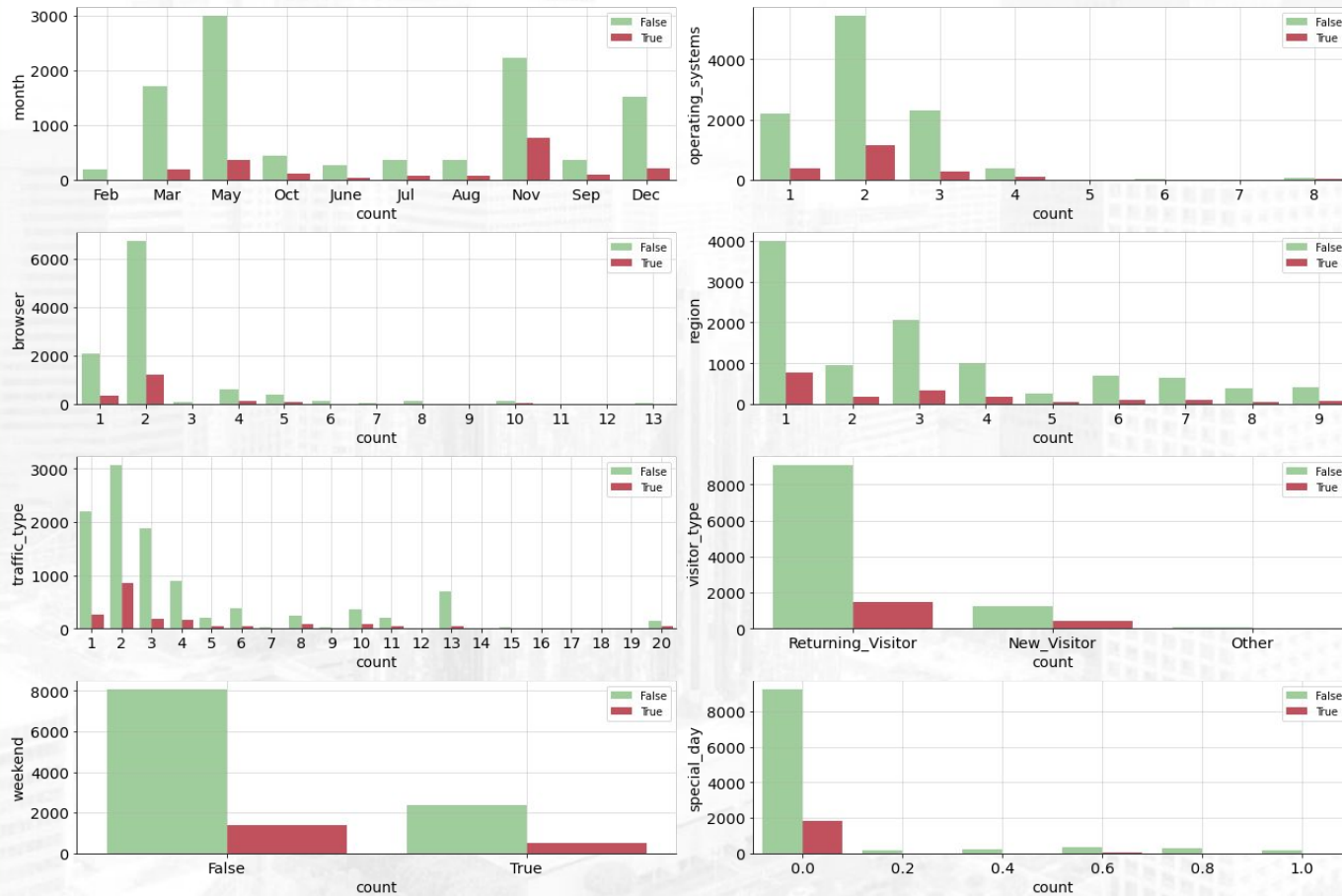
3. Multivariate Analysis (Kolom Numerik)



- Pengunjung website mengunjungi halaman web product related lebih banyak dibandingkan dengan halaman administratif dan information.
- Pengunjung yang menghasilkan revenue cenderung lebih lama mengakses halaman web product related.
- Pengunjung dengan bounce rates yang lebih rendah, exit rates yang lebih rendah, dan page values yang lebih tinggi cenderung menghasilkan revenue.

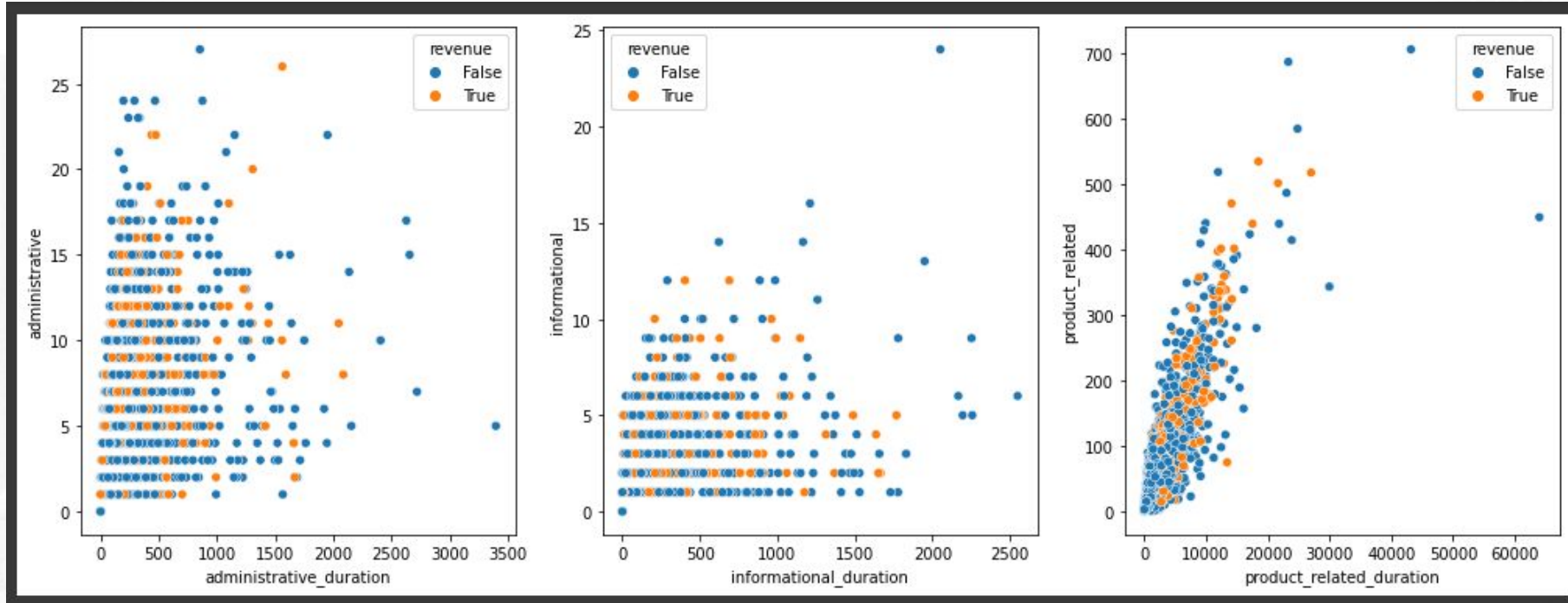
Exploratory Data Analysis

3. Multivariate Analysis (Feature Kategorikal)



- Jumlah pengunjung website yang menghasilkan revenue terbanyak pada bulan November dan terendah pada bulan Februari dan Juni.
- Pengunjung website yang menghasilkan revenue mayoritas menggunakan browser dan OS tipe 2.
- Sumber traffic yang paling banyak menghasilkan revenue adalah traffic type 2.
- Returning visitor banyak yang tidak melakukan transaksi yang menghasilkan revenue.
- Pengunjung website lebih banyak bertransaksi di weekday daripada di weekend.

Exploratory Data Analysis



- administrative dan administrative_duration, informational dan informational_duration, product_related dan product_related_duration termasuk korelasi positif.
- meskipun pengunjung menghabiskan banyak waktu pada halaman product_related, tapi menghasilkan pendapatan yang rendah. Ada outliers tertentu yang menghabiskan waktu 3000 detik (sekitar 8 jam) tapi tidak melakukan transaksi apapun

Exploratory Data Analysis

Kesimpulan Multivariate Analysis

A. Bagaimana korelasi antara masing-masing feature dan label. Kira-kira feature mana saja yang paling relevan dan harus dipertahankan?

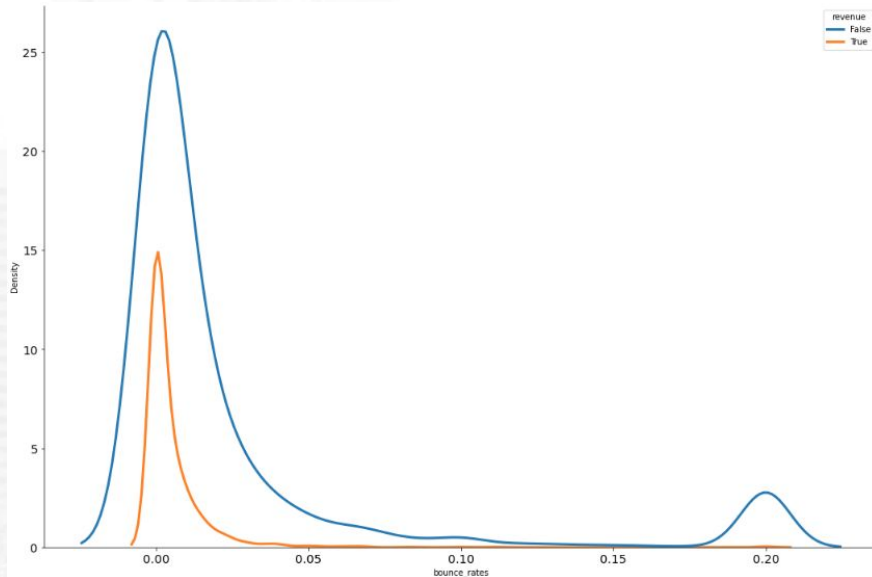
- Berdasarkan analisis korelasi semua kolom numerik dan kategorikal memiliki korelasi ke target kecuali feature region. Berikut ini adalah 3 feature dari masing-masing tipe kolom yang paling tinggi nilai korelasi terhadap target.
 - Numerik: page_values, exit_rates, dan product_related_duration
 - Kategorikal: month, traffic_type, dan visitor_type

B. Bagaimana korelasi antar-feature, apakah ada pola yang menarik? Apa yang perlu dilakukan terhadap feature itu?

- Berdasarkan spearman's correlation yang ditunjukkan oleh heatmap, ada korelasi antar feature numerik yang significant di atas 0.8 yaitu kolom jumlah page dan jumlah durasi page memiliki korelasi yang tinggi yaitu 0.88 - 0.95.
- Karena terdapat korelasi tinggi antar feature numerik dan ada feature kategorikal yang tidak berkorelasi ke label maka
 - Salah satu dari pasangan kolom numerik yang berkorelasi tinggi harus didrop
 - kolom region bisa didrop

Business Insights

Bounce Rates



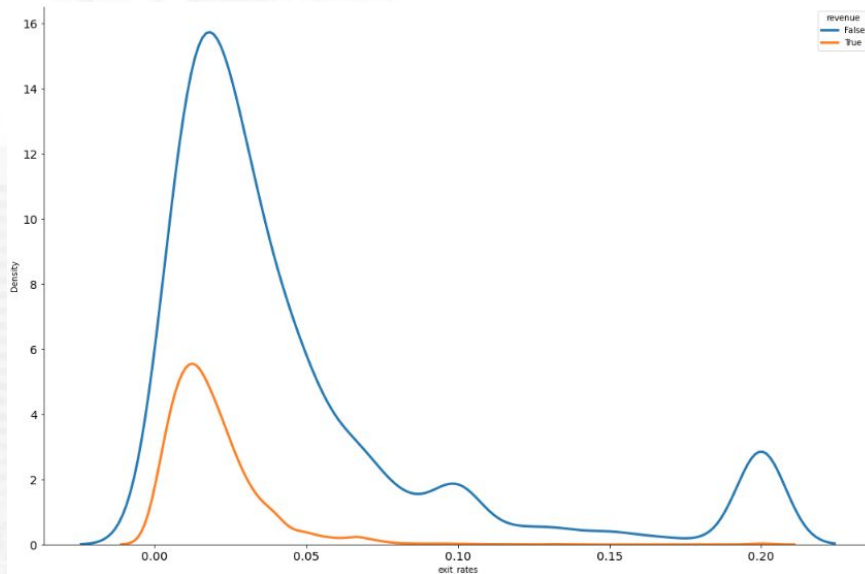
Bounce rate yang tinggi pada halaman ecommerce, secara umum akan menghasilkan revenue yang sedikit. Untuk ecommerce, bounce rate dibawah 20% atau diatas 90% dapat menjadi tanda bahwa ada hal yang tidak beres. Ketika bounce rate terlalu rendah, bisa saja disebabkan masalah pada analytics setup. Sementara bounce rate yang terlalu tinggi, bisa jadi pertanda bahwa site ecommerce tidak memiliki performance yg baik. [Reference](#)

Pengunjung yang menghasilkan revenue cenderung memiliki bounce rate yang lebih rendah dibandingkan yang tidak menghasilkan revenue yang mana sesuai dengan reference dan teori secara umum. Namun, apabila dilihat secara keseluruhan, rata-rata dan median bounce rate rendah: 0.3-2.3% dan nilai max 20% sedangkan persentase jumlah pengunjung yang menghasilkan revenue sangat kecil sekitar 15% dari seluruh total pengunjung.

Rekomendasi : Lakukan pengecekan terhadap analytics setup, apakah ada kesalahan setting sehingga ada bounce rate yang tidak terdeteksi.

Business Insights

Exit Rates



Exit Rates adalah persentase ketika visitor keluar dari website kita dari page/halaman tertentu (halaman terakhir sebelum visitor meninggalkan website kita). Exit rate yang tinggi pada sebuah halaman spesifik dapat menjadi pertanda ada yg salah dengan halaman tersebut terutama high exit rate pada halaman check-out atau halaman produk. Tetapi, ada pengecualian pada halaman-halaman tertentu, Sebagai contoh pada Thank you Page, setelah transaksi, exit rate tinggi bisa jadi termasuk hal yang normal.

[Reference](#)

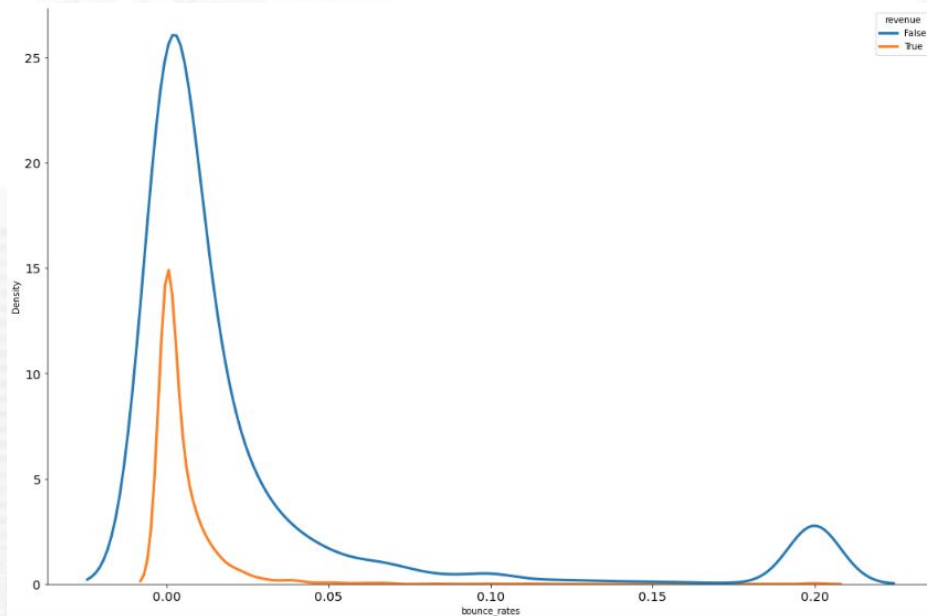
Exit rate secara langsung terkait dengan bounce rate, karena parameter bagi keduanya adalah ketika visitor meninggalkan website [Reference](#). Benchmark hampir sama dengan bounce rate yang mana dibawah 20% terlalu rendah dan diatas 90% terlalu tinggi.

Jumlah pengunjung yang menghasilkan revenue cenderung memiliki exit rate yang lebih rendah dibandingkan yang tidak menghasilkan revenue yang mana sesuai dengan reference dan teori secara umum. Rata-rata dan median exit rate rendah: 2.5-4% , max values nya juga 20%.

Rekomendasi : lakukan pengecekan apakah analytics setup sudah benar.

Business Insights

Product Related



Secara general, visitor yang mengunjungi pages bertipe product related memiliki intensi membeli yang lebih tinggi atau sedang dalam tahap membuat decision untuk membeli. Product page yang menghasilkan hanya sedikit revenue dapat mengindikasikan bahwa product page tersebut tidak mendemonstrasikan produk dengan baik atau secara interface juga tidak cukup membuat user betah dipage tersebut.

[Reference](#)

Diantara semua jenis page (product page, administrative, informational), product page adalah yang jumlahnya paling banyak dan paling lama dikunjungi oleh visitor. (90% dari total kunjungan dan durasi kunjungan semua page).

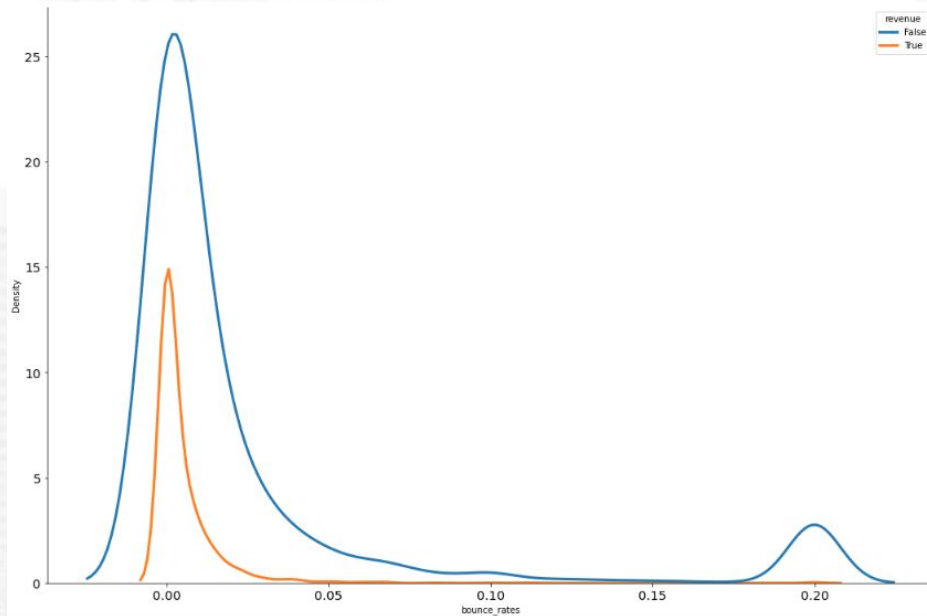
Bila dilihat pun dari seluruh transaksi dengan revenue, sebagian besar adalah berasal dari visit ke product page.

Tetapi 90% total kunjungan page dibandingkan dengan 15% sesi yang menghasilkan conversion, terbilang rendah.

Rekomendasi : Lakukan pengecekan terhadap product related page dari segi interface, kemudahan akses, kejelasan informasi dan demonstrasi produk, dll. Apakah high product page visit tapi low conversion disebabkan oleh hal-hal tersebut.

Business Insights

Page Values



Nilai page value yang tinggi menunjukkan page mana yang lebih attractive dimata visitor, dengan mengetahui page yang memiliki high value tinggi, kita bisa memusatkan page tersebut untuk generate lebih banyak revenue (reference). Sesi yang menghasilkan pendapatan cenderung memiliki page value yang lebih tinggi dibandingkan yang tidak menghasilkan revenue yang mana sesuai dengan reference dan teori secara umum. [Reference](#)

Apabila dilihat secara keseluruhan, nilai page value sangat skewed. Hampir seluruh nilainya sangat kecil yang apabila dilihat dari statistik deskriptif, hingga 75% data masi bernilai 0 yang berarti banyak sekali page yang tidak cukup 'menarik' bagi visitor.

Rekomendasi : Memperbaiki kualitas konten di website agar lebih menarik.

Kontribusi

Pembagian tugas di stage 1: Seluruh anggota melakukan EDA, tapi masing-masing dengan feature yang berbeda.

Nama: M. Irfan F □ Administrative, Administrative_Duration, Informational, Informational_Duration

Nama: Ramado Dipradelana I □ BounceRates, ExitRates, PageValues

Nama: Ni Putu Tasya □ TrafficType, VisitorType, ProductRelated, ProductRelated_Duration

Nama: Bima Purnomo Sandi □ SpecialDay, Month, Weekend

Nama: Deni Indra Permana □ OperatingSystems, Browser, Region

Business Insights (Lanjutan)

Dari data yang tersedia, secara teori sebenarnya hampir semua variabel bisa memiliki pengaruh terhadap transaksi yang menghasilkan revenue, even dari operating systemnya pun bisa. Tapi sayangnya beberapa variabel bentuknya masih berupa binary, dan gak ada tabel acuannya (contoh: TrafficType, etc) sehingga tidak bisa semuanya dijadikan insight yang mendetail.

Bounce Rate

Teori	Dataset & EDA
<p>Bounce rate merupakan salah satu metrics yang menunjukkan seberapa baik performa sebuah ecommerce site. Sehingga bounce rate yang tinggi pada halaman ecommerce, secara umum akan menghasilkan revenue yang sedikit. (reference)</p> <p>Untuk ecommerce, bounce rate dibawah 20% atau diatas 90% dapat menjadi tanda bahwa ada hal yang tidak beres. Ketika bounce rate terlalu rendah, bisa saja disebabkan masalah pada analytics setup. Sementara bounce rate yang terlalu tinggi, bisa jadi pertanda bahwa site ecommerce tidak memiliki performance yg baik (reference)</p>	<p>Sesi yang menghasilkan pendapatan cenderung memiliki bounce rate yang lebih rendah dibandingkan yang tidak menghasilkan revenue. (ini sesuai dengan reference dan teori secara umum)</p> <p>Namun apabila dilihat secara keseluruhan, rata-rata dan median bounce rate rendah: 0.3% - 2.3% . dan nilai max 20%. Tetapi persentasi transaksi yang melakukan revenue sangat kecil sekitar 15% dari seluruh total sesi.</p> <p>Rekomendasi: Lakukan pengecekan terhadap analytics setup, apakah ada kesalahan setting sehingga ada bounce rate yang tidak terdeteksi.</p>

Business Insights (Lanjutan)

Product Related Page and Revenue

Teori	Dataset & EDA
<p>Secara general, visitor yang mengunjungi pages bertipe product related memiliki intensi membeli yang lebih tinggi atau sedang dalam tahap membuat decision untuk membeli.</p> <p>Product page yang yang menghasilkan hanya sedikit transaksi (revenue) dapat mengindikasikan bahwa product page tersebut tidak mendemonstrasikan produk dengan baik atau secara interface juga tidak cukup membuat user betah dipage tersebut. (reference)</p>	<p>Diantara semua jenis page (product page, administrative, informational), product page adalah yang jumlahnya paling banyak dan paling lama dikunjungi oleh visitor. (90% dari total kunjungan dan durasi kunjungan semua page).</p> <p>Bila dilihat pun dari seluruh transaksi dengan revenue, sebagian besar adalah berasal dari visit ke product page. Tetapi 90% total kunjungan page dibandingkan dengan 15% sesi yang menghasilkan conversion, terbilang rendah.</p> <p>Rekomendasi: Lakukan pengecekan terhadap product related page dari segi interface, kemudahan akses, kejelasan informasi dan demonstrasi produk, dll. Apakah high product page visit tapi low conversion disebabkan oleh hal-hal tersebut.</p>

Business Insights (Lanjutan)

Exit Rates

Teori	Dataset & EDA
<p>Exit Rates adalah persentase ketika visitor keluar dari website kita dari page/halaman tertentu (halaman terakhir sebelum visitor meninggalkan website kita). Exit rate yang tinggi pada sebuah halaman spesifik dapat menjadi pertanda ada yg salah dengan halaman tersebut terutama high exit rate pada halaman check-out atau halaman produk. Tetapi ada pengecualian pada halaman-halaman tertentu, Sebagai contoh pada Thank you Page, setelah transaksi, exit rate tinggi bisa jadi termasuk hal yang normal. (reference)</p> <p>Exit rate secara langsung terkait dengan bounce rate, karena parameter bagi keduanya adalah ketika visitor meninggalkan website (reference).</p> <p>Benchmark hampir sama dengan bounce rate 20%-90% Dibawah 20% terlalu rendah, 90% terlalu tinggi</p>	<p>Sesi yang menghasilkan pendapatan cenderung memiliki exit rate yang lebih rendah dibandingkan yang tidak menghasilkan revenue. (ini sesuai dengan reference dan teori secara umum)</p> <p>Rekomendasi:</p> <ul style="list-style-type: none"> - Rata-rata dan median exit rate rendah: 2.5% - 4% , max values nya juga 20%. lakukan pengecekan apakah analytics setup sudah benar.

Business Insights (Lanjutan)

Page Values

Teori	Dataset & EDA
<p>Nilai page value yang tinggi menunjukkan page mana yang lebih attractive dimata visitor, dengan mengetahui page yang memiliki high value tinggi, kita bisa memusatkan page tersebut untuk generate lebih banyak revenue. (reference)</p>	<p>Sesi yang menghasilkan pendapatan cenderung memiliki page value yang lebih tinggi dibandingkan yang tidak menghasilkan revenue. (ini sesuai dengan reference dan teori secara umum)</p> <p>Apabila dilihat secara keseluruhan, nilai page value sangat skewed. Mostly nilainya sangat kecil (apabila dilihat dari statistik deskriptif, Q1-Q3 nya 0) yang berarti banyak sekali page yang tidak cukup ‘menarik’ bagi visitor.</p>

	count	mean	std	min	25%	50%	75%	max
page_values	9864.0	5.802	18.104	0.0	0.000	0.000	0.000	361.764