



USER RETENTION ANALYST

ANALISA PERILAKU KONSUMEN



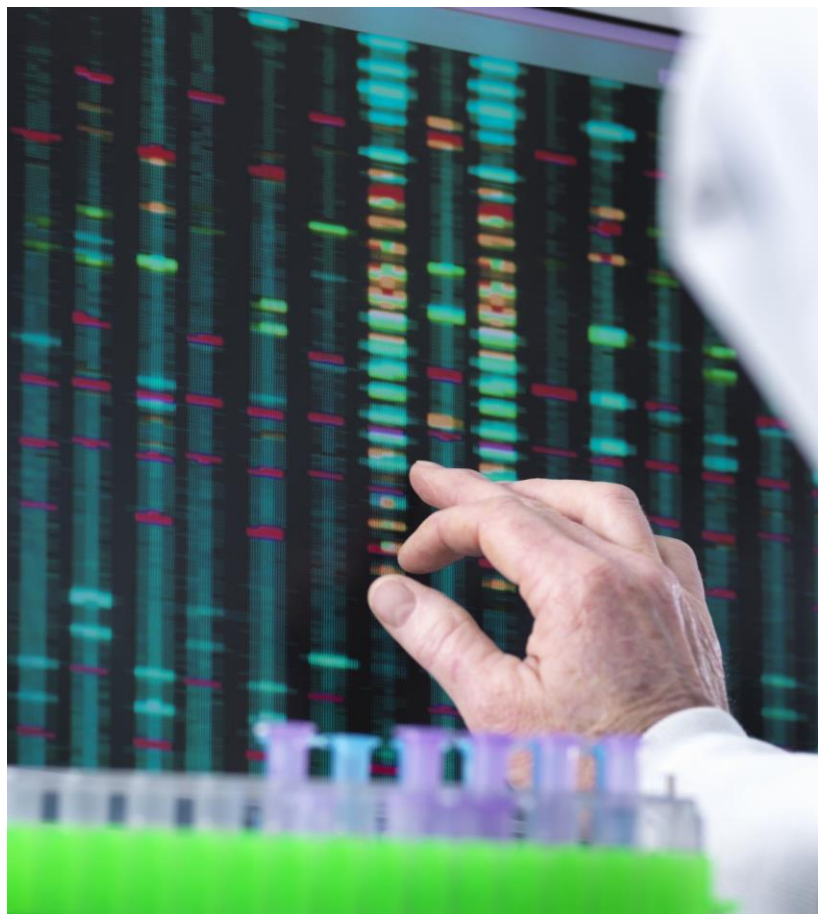
PENGERTIAN USER RETENTION ANALYSIS

- *User retention analysis* merupakan suatu kegiatan untuk mengetahui apakah terdapat perilaku *user* atau pengguna yang kembali bertransaksi atau menggunakan suatu produk (*retained*). Dengan proses ini, kita dapat menganalisa sejauh mana *user* terus menggunakan suatu produk dalam jangka waktu tertentu.

COHORT ANALYSIS

- *Cohort* merupakan suatu kelompok yang mempunyai kesamaan, seperti tanggal pendaftaran, bulan pembelian, dan lokasi geografis. *Cohort analysis* merupakan proses mengidentifikasi kelompok *user* atau pengguna ini dari waktu ke waktu sehingga kita dapat mengidentifikasi pola atau perilaku umum *user* atau pengguna.





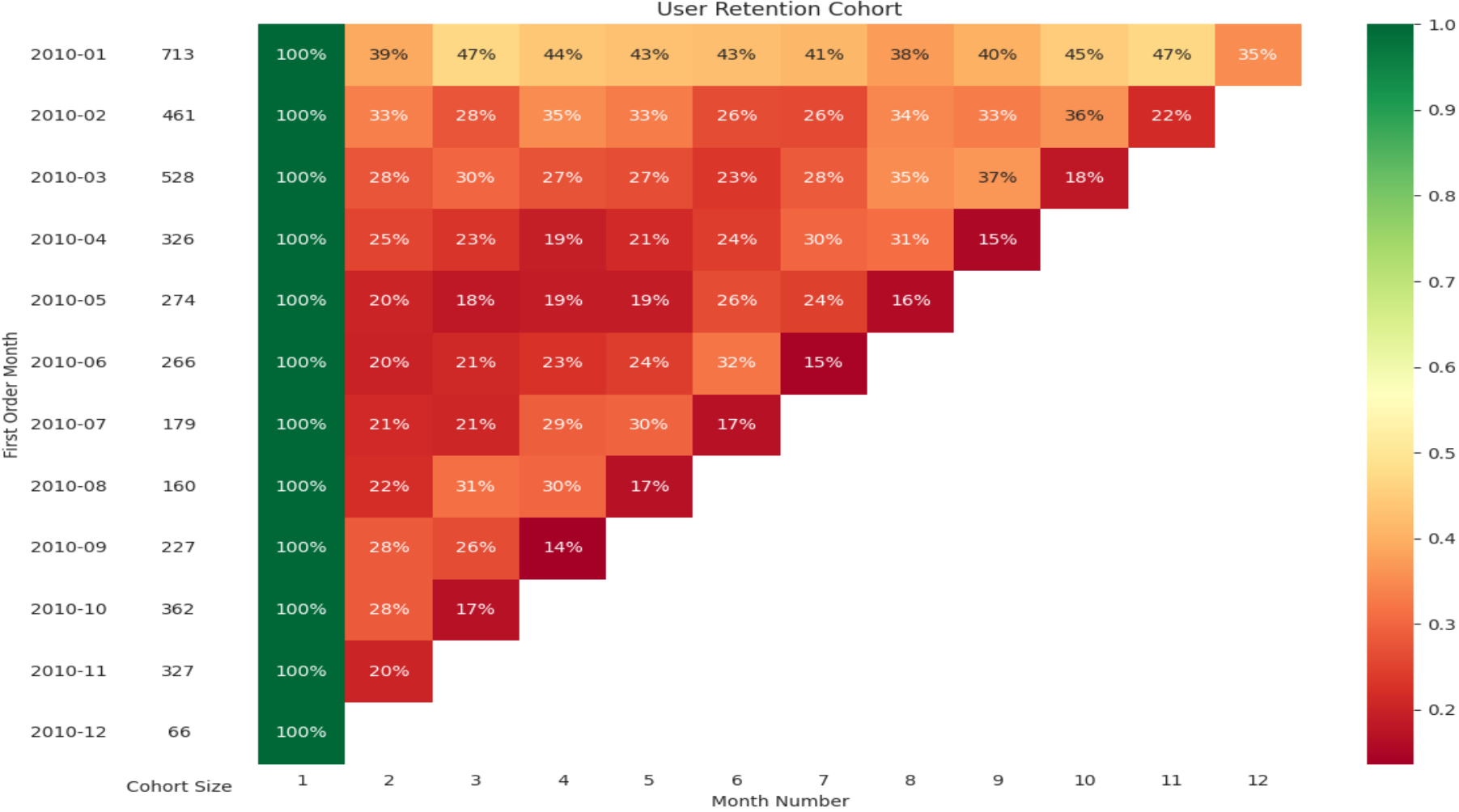
CREATING A USER RETENTION COHORT IN PYTHON

1. Impor data transaksi ke DataFrame
`df = pd.read_csv ('Online Retail Data.csv ', header=0)`
2. Lakukan data cleansing, buat kolom tahun dan bulan (untuk melihat pola transaksi bulanan)
`df_clean = df.copy ()`
`df_clean ['year month'] = df_clean ['order date'].dt.to_period ('M ')`
`... # data cleansing`
3. Agregat data transaksi ke bentuk summary total transaksi/ order setiap pengguna setiap bulan
`df_user_monthly = df_clean.groupby (['customer_id','year_month'], as_index=False)`
`.agg (order_cnt= ('order_id','nunique'))`
4. Buat kolom sebagai cohort dari pengguna, misal cohort bulan pertama kali bertransaksi
`df_user_monthly ['cohort'] = df_user_monthly.groupby ('customer id')['year month`
`'].transform('min '`

5. Hitung jarak bulan antara bulan transaksi (poin 2) dengan bulan pertama kali transaksi (poin 4) dan jumlahkan dengan 1 agar jarak bulan 0 menjadi 1 yang berarti bulan pertama, dst
`from operator import attrgetter
df_user_monthly ['period_num'] = (df_user_monthly ['year
month'] -
df_user_monthly ['cohort']).apply (attrgetter ('n')) + 1`
6. Buat tabel pivot dengan index berupa cohort (poin 4), kolom berupa jarak bulan (poin 5), dan nilainya adalah banyaknya pengguna unik (count unique dari ID pengguna)
`df_cohort_pivot = pd.pivot_table (df_user_monthly, index=
'cohort', columns= 'period_num',
values= 'customer_id', aggfunc=pd .Series.nunique)`
7. Hitung banyaknya pengguna di masing-masing cohort (pengguna yang pertama kali transaksi di bulan tersebut) dan bagi semua nilai di tabel pivot tadi dengan nilai tersebut sesuai baris index-nya sebagai retention rate
`cohort_size = df_cohort_pivot. iloc [; ,0]
df_retention_cohort = df_cohort_pivot.divide (cohort_size,
axis=0)`
8. Tampilkan tabel pivot yang sudah berisi nilai *retention rate* tersebut dalam *heatmap*.



TABEL COHORT



HASIL ANALISA

- Pengguna **paling banyak pertama kali bertransaksi pada Januari 2010** (713 pengguna)
- Cohort pengguna tersebut juga yang **paling banyak bertransaksi kembali di bulan ke 2 (39% retention rate)** dibanding cohort yang lain
- Selain itu **Cohort tersebut yang paling loyal bertransaksi selama bulan – bulan berikutnya dengan retention rate ~40%+**
- Sayangnya, **sebagian besar pengguna tidak kembali bertransaksi** , terlihat dari retention rate di banyak Cohort dan bulan yang nilai tak sampai 50%
- Yang cukup mengkhawatirkan , **retention rate di Desember 2010 menjadi terendah untuk semua Cohort** pengguna di banding bulan – bulan sebelumnya.

~ SEKIAN DAN TERIMAKASIH ~