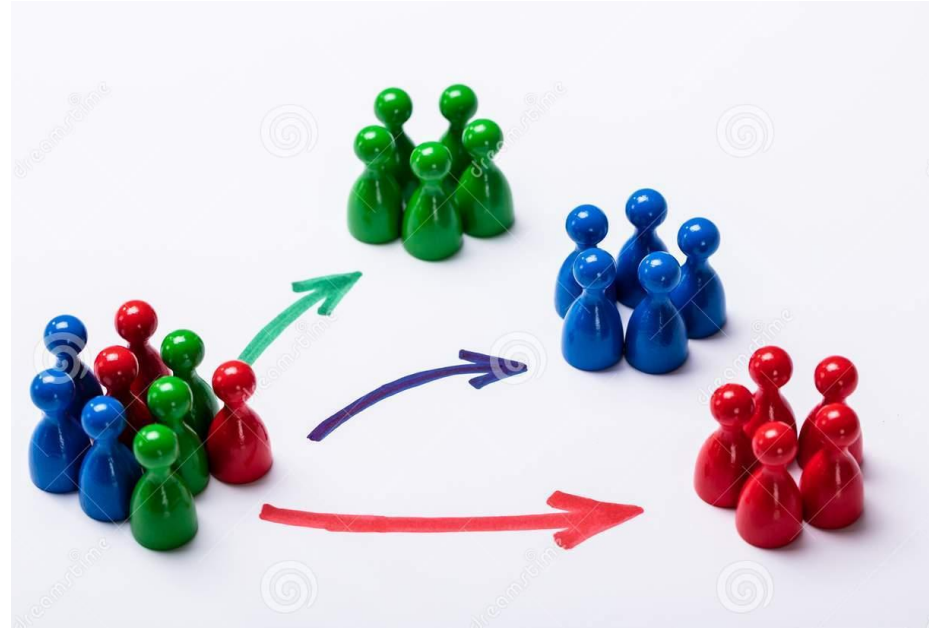


Customer Segmentasi dengan Algoritma K-Prototype



The Problem

Perusahaan sedang berusaha untuk mengenal lebih baik pelanggannya. Dengan tujuannya agar perusahaan dapat membuat strategi pemasaran yang lebih tepat dan efisien bagi tiap-tiap pelanggan.

Salah satu teknik yang bisa dilakukan untuk mengenal lebih baik pelanggan adalah dengan melakukan segmentasi pelanggan. Yaitu dengan mengelompokkan pelanggan-pelanggan yang ada berdasarkan kesamaan karakter dari pelanggan tersebut. Untuk melakukan hal tersebut Kita akan menggunakan teknik unsupervised machine learning.

Salah satunya teknik yang dapat digunakan adalah pengaplikasian machine learning menggunakan algoritma K-Prototypes. Algoritma K-Prototypes merupakan gabungan dari K-Means dan juga K-Modes yang dapat digunakan untuk melakukan segmentasi dengan data.

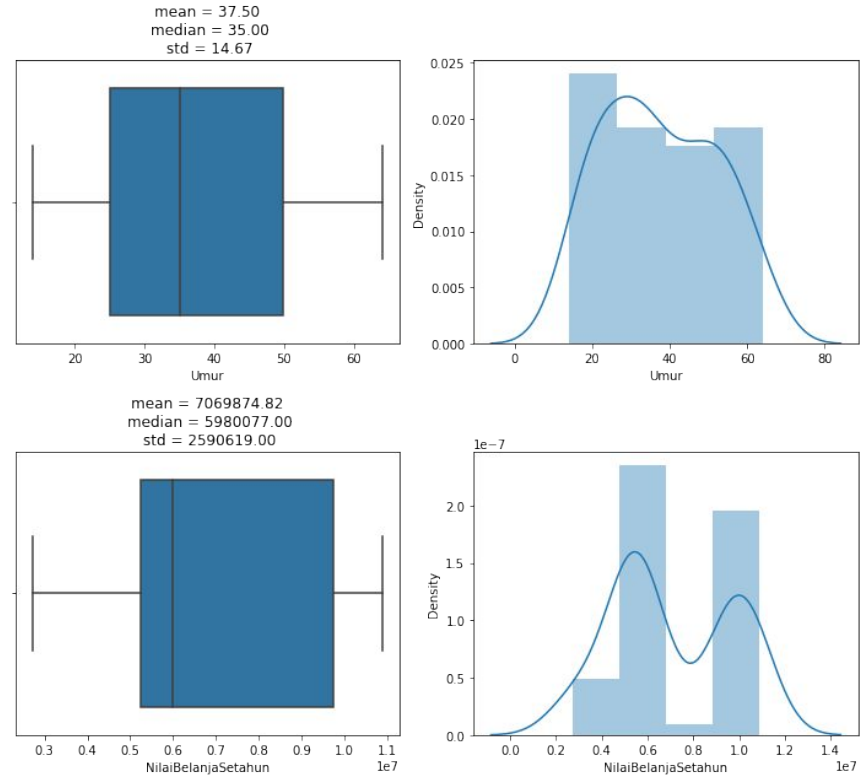
Dataset

Dataset yang kita gunakan memiliki tujuh kolom dengan penjelasan sebagai berikut:

- Customer ID : Kode pelanggan dengan format campuran teks CUST- diikuti angka.
- Nama Pelanggan : Nama dari pelanggan dengan format teks tentunya.
- Jenis Kelamin : Jenis kelamin dari pelanggan, hanya terdapat dua isi data kategori yaitu Pria dan Wanita.
- Umur : Umur dari pelanggan dalam format angka.
- Profesi : Profesi dari pelanggan, juga bertipe teks kategori yang terdiri dari Wiraswasta, Pelajar, Professional, Ibu Rumah Tangga, dan Mahasiswa.
- Tipe Residen : Tipe tempat tinggal dari pelanggan kita, untuk dataset ini hanya ada dua kategori: Cluster dan Sector.
- Nilai Belanja Setahun : Merupakan total belanja yang sudah dikeluarkan oleh pelanggan tersebut.

Eksplorasi Data

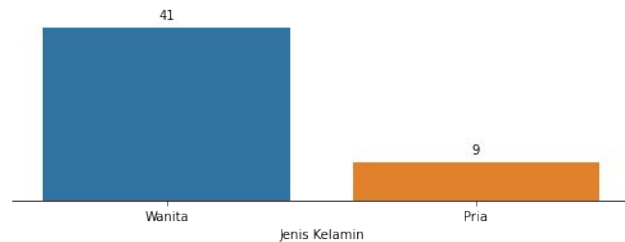
- Rata-rata dari umur pelanggan adalah 37.5 tahun
- Rata-rata dari nilai belanja setahun pelanggan adalah 7,069,874.82



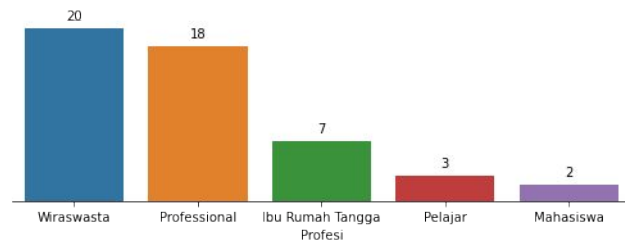
Eksplorasi Data

- Jenis kelamin pelanggan di dominasi oleh wanita sebanyak 41 orang (82%) dan laki-laki sebanyak 9 orang (18%)
- Profesi terbanyak adalah Wiraswasta (40%) diikuti dengan Professional (36%) dan lainnya sebanyak (24%)
- Dari seluruh pelanggan 64% dari mereka tinggal di Cluster dan 36% nya tinggal di Sektor

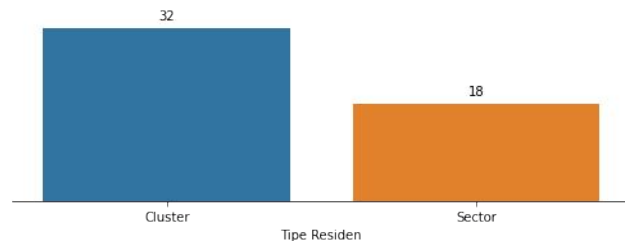
Count Plot Jenis Kelamin



Count Plot Profesi



Count Plot Tipe Residen



Feature Engineering



Standarisasi Kolom Numerik

- Kolom Umur
- NilaiBelanjaSetahun



Konversi Kategorikal Data dengan Label Encoder

- Jenis Kelamin
- Profesi
- Tipe Residen

Clustering dan Algoritma K-Prototypes

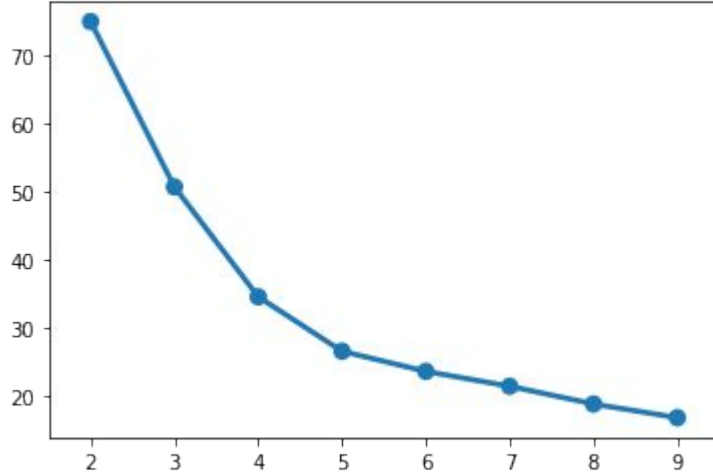
Clustering adalah proses pembagian objek-objek ke dalam beberapa kelompok (cluster) berdasarkan tingkat kemiripan antara satu objek dengan yang lain.

Terdapat beberapa algoritma untuk melakukan clustering ini. Salah satu yang populer adalah k-means.

K-means biasanya hanya digunakan untuk data-data yang bersifat numerik. Sedangkan untuk yang bersifat kategorikal saja, kita bisa menggunakan k-modes.

Lalu bagaimana apabila di data kita terdapat gabungan kategorikal dan numerikal variabel? Jawabannya kita bisa menggunakan algoritma k-prototype yang merupakan gabungan dari k-means dan k-modes.

Mencari Jumlah Cluster yang Optimal



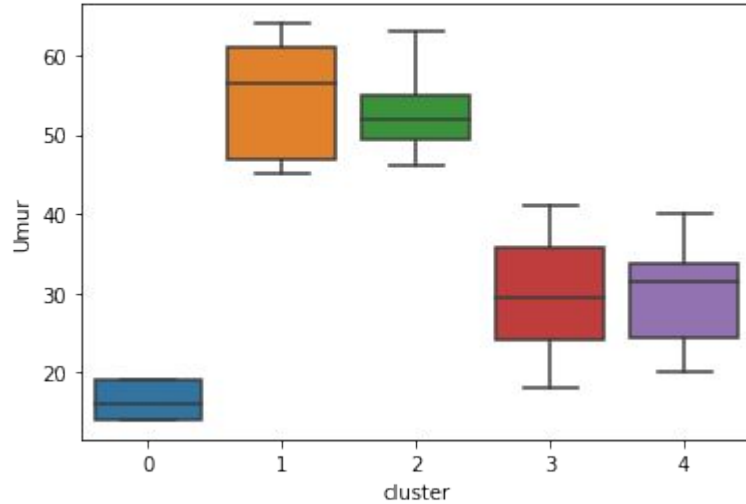
5

Titik siku dari plot adalah pada saat $k = 5$.

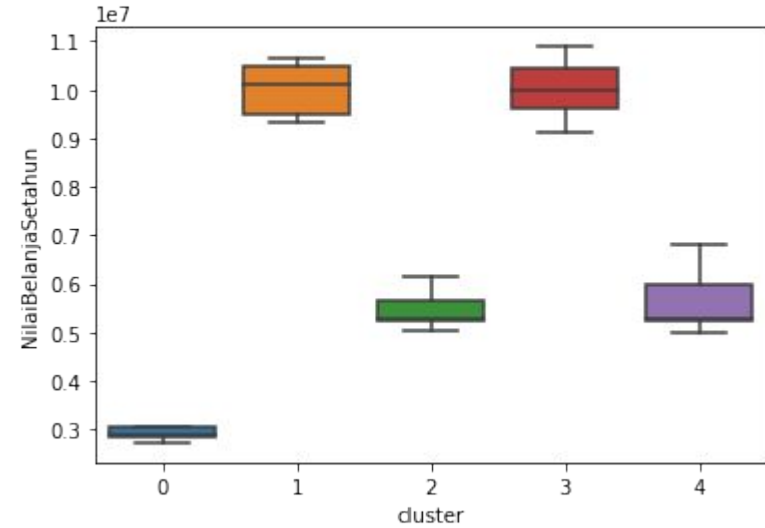
Sehingga kita memutuskan untuk menggunakan 5 sebagai jumlah cluster optimalnya.

Hasil Data Clustering

Box Plot Umur

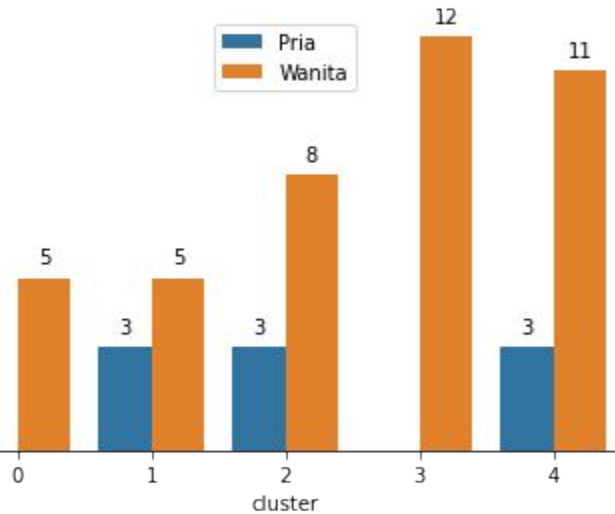


Box Plot NilaiBelanjaSetahun

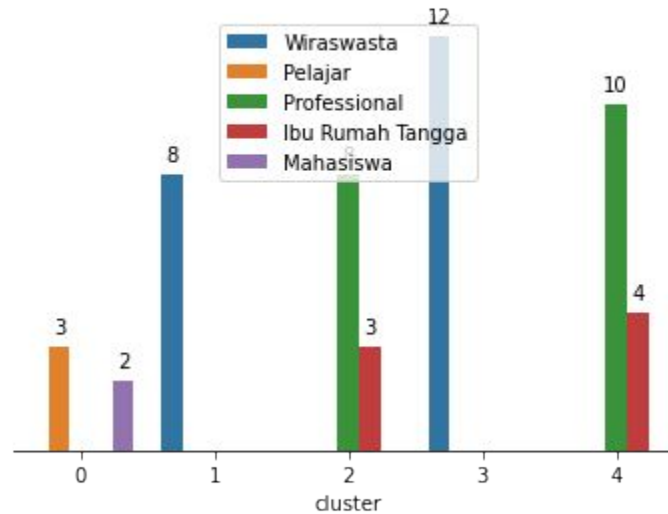


Hasil Data Clustering

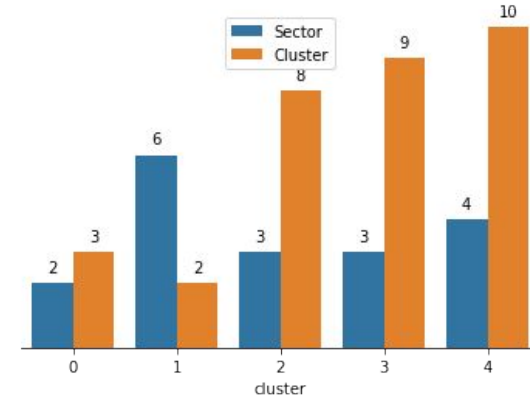
Count Plot Jenis Kelamin



Count Plot Profesi



Count Plot Tipe Residen



5 Cluster (segmen) Customer hasil Clustering K-Prototype:

- **Diamond Young Entrepreneur:** Wiraswasta rata-rata nilai transaksi mendekati 10 juta, umur 18 - 41 tahun
- **Diamond Senior Entrepreneur:** Wiraswasta rata-rata nilai transaksi mendekati 10 juta, umur 45 - 64 tahun
- **Silver Students:** Pelajar dan Mahasiswa rata-rata umur 16 tahun dan nilai belanja setahun 3 mendekati juta.
- **Gold Young Member:** Profesional dan Ibu rumah tangga , umur sekitar 20 - 40 tahun dan nilai belanja setahunnya mendekati 6 juta.
- **Gold Senior Member:** Profesional dan Ibu rumah tangga umur 46 - 63 tahun dan nilai belanja setahunnya mendekati 6 juta.

CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Nur Purwanto

Has Completed in

Data Science in Marketing : Customer Segmentation with Python - Part 1

Jun 15, 2022



CERTIFICATE OF COMPLETION

This certificate is proudly presented to

Nur Purwanto

Has Completed in

Data Science in Marketing : Customer Segmentation with Python - Part 2

Jun 15, 2022



THANK YOU!!!!!!

Mari Saling Terhubung

[[Linkedin](#)] [[Github](#)] [[Medium](#)] [[Site](#)]

[[Code](#)]