

# AI Mastery Course

## Module 3

Data analytics with python- applied analytics


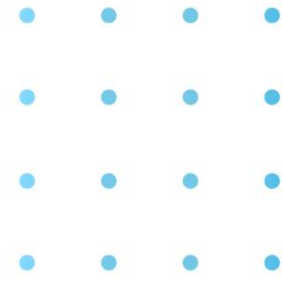
### Section

Exploratory Data Analytics (EDA)



# Learning Objectives

Di akhir modul ini, Anda akan dapat:

- Menjelaskan derived variables/ variabel turunan dan metode yang digunakan untuk menganalisanya
  - Memahami overview dari exploratory data analysis (EDA) dan manfaatnya
  - Mengidentifikasi perbedaan metode yang tersedia di EDA dan 1 contohnya
- 
- 



# Agenda

01

## VARIABLES

- Derived Variables

02

## EDA

- Basic Exploratory Data Analysis
- Methods for EDA and Examples

03

## CONCLUSION

- Summary
- Quiz



01

## VARIABLES

- Derived Variables

# Derived Variables/Variabel Turunan

**Di bawah ini adalah beberapa poin penting dari variable turunan:**

- Dalam konteks statistical analysis, variable turunan merupakan variabel yang diturunkan dari dua atau lebih variabel utama
- Variabel turunan dibentuk dari melakukan kalkulasi atau kategorisasi dari variabel yang ada di dataset
- Beberapa contoh dari variabel turunan adalah rasio, persentase, index dan rate
- Variabel turunan berdasarkan pada variabel utama, maka perubahan yang terjadi pada variabel utama akan memiliki dampak pada variabel turunan
- Variable turunan bisa memiliki beberapa properti yang tidak diharapkan dan memerlukan metode penanganan yang berbeda

# Analysis of Derived Variables

## Beberapa hal terkait analisa dari variabel turunan.

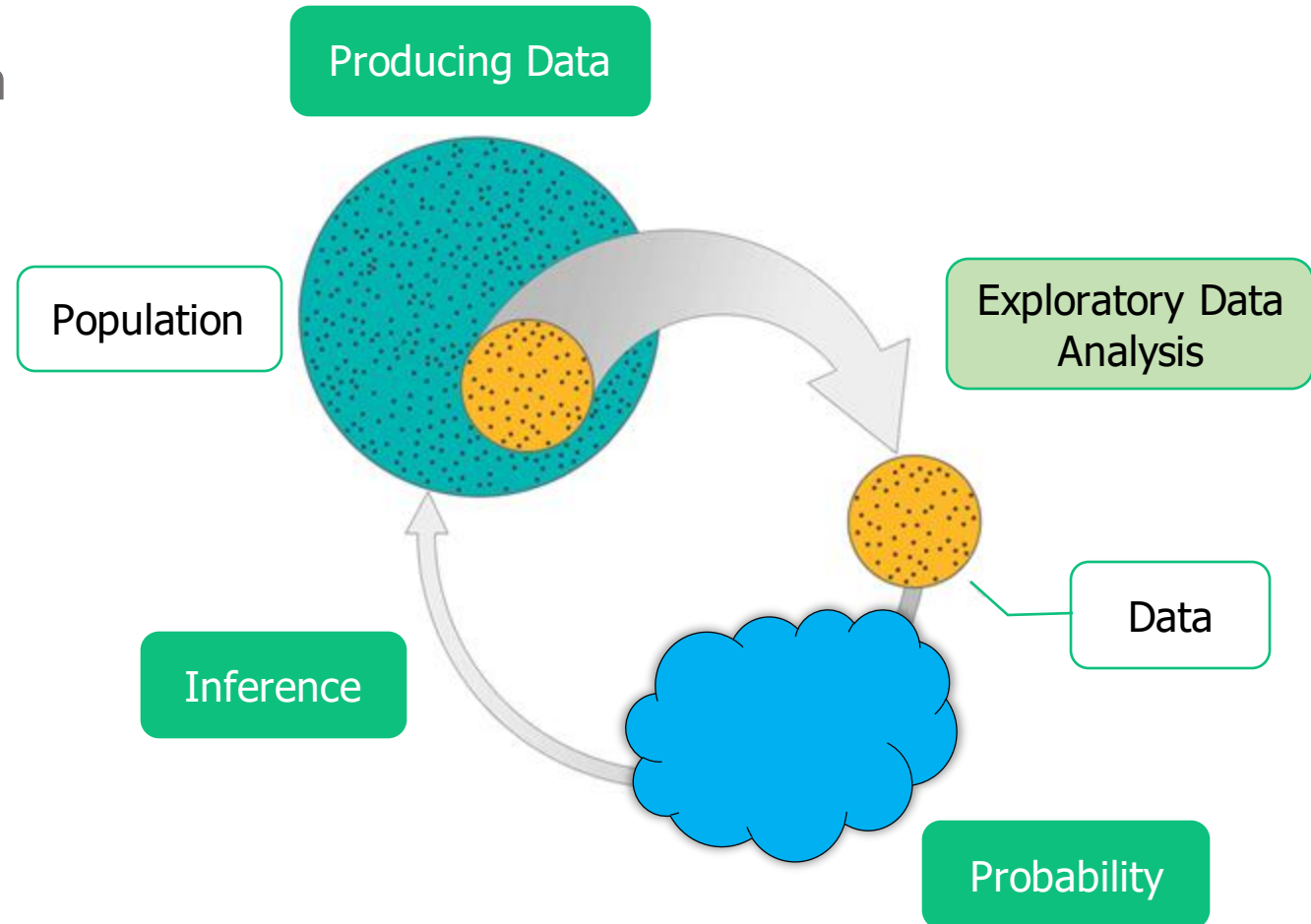
- Analisa dari variabel turunan memuat beberapa metode untuk mengambil sekumpulan pengukuran dan mengkombinasikannya menjadi fitur ringkas yang penting
- Metode statistik digunakan untuk membentuk model atau prediksi, menghasilkan nilai yang diprediksi dan residual
- Nilai yang diprediksi adalah nilai yang diprediksi oleh model, dimana untuk kebanyakan kasus merupakan single variable. Kita bisa menganalisa variabel berikut juga relasinya terhadap variabel lain di model tersebut atau di luar model tersebut
- Residual merupakan hal dimana model tidak dapat prediksi. Residual digunakan untuk diagnosa model dan mengidentifikasi masalah yang terkait dengan model tersebut
- Pada praktiknya, variabel turunan dibuat selama melakukan analisa statistik, dengan banyak variasi, yang tergantung pada teknik analisis yang digunakan
- Ada banyak tools yang tersedia untuk menganalisa variabel turunan

## 02 EDA

- Basic Exploratory Data Analysis
- Methods for EDA and Examples

# Exploratory Data Analysis

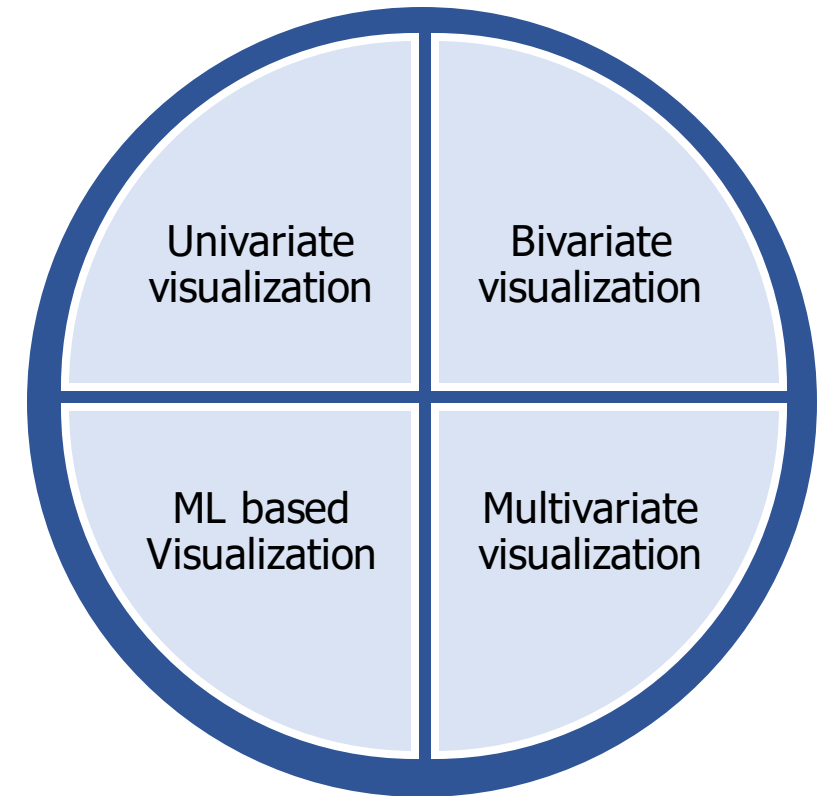
- Exploratory data analysis (EDA) dilakukan pada sebuah data sebagai serangkaian upaya investigasi awal, untuk menemukan suatu pola, mendeteksi anomali, hypothesis testing dan untuk memvalidasi asumsi menggunakan ringkasan statistik dan representasi visual
- EDA digunakan untuk memastikan bahwa data yang dimiliki siap untuk digunakan oleh algoritma machine learning, dan membantu untuk menentukan algoritma yang tepat dari dataset yang ada
- EDA juga digunakan untuk menentukan feature variable yang dapat digunakan oleh machine learning





# Methods of Exploratory Data Analysis

- Ada beberapa teknik eksplorasi yang tersedia untuk melakukan EDA dan hasilnya bisa dikomparasikan
- Setelah melakukan analisa data dan memahaminya, kita bisa melakukan data collection dan data cleaning untuk mentransformasi data agar sesuai dengan business requirements



# EDA with Data Visualization

- Univariate plot biasa digunakan untuk menampilkan distribusi data dengan satu variabel. Variabel tersebut bisa berbentuk kategori (mis: jenis kelamin, ras, negara, kota, dll) atau berbentuk kuantitatif (mis: umur, berat badan, inflasi, penilaian, dll).
- Data Bivariate menampilkan hubungan antara dua variabel. Tipe grafik ini tergantung level pengukuran dari variabel (kategori atau kuantitatif)
- Data Multivariate menampilkan hubungan antara tiga atau lebih variabel.
- Dimensionality reduction digunakan untuk memahami fields dari data yang menjelaskan paling banyak varians antara data observasi dan kemungkinan volume data yang dikurangi
- Pola perilaku dapat lebih mudah diidentifikasi dengan melakukan pengelompokan (clustering) data observasi yang serupa pada dataset ke dalam pengelompokan yang berbeda (dengan cara mengelompokkan data menjadi beberapa kumpulan data kecil)

# Approaches and Techniques

## Categorical V/s Continuous

1. Probability Distribution analysis - using distplot
2. Swarm plot and boxplot
3. ANOVA - f\_classif

## Continuous v/s continuous

1. Scatterplot
2. Correlation Analysis
3. ANOVA - f\_regress

## Categorical v/s Categorical

1. Barplot - of ratio of frequency/count in multiple groups
2. Chi Square Test
3. Histogram/countplot



## 03

## CONCLUSION

- Summary
- Quiz

# Summary

1. Variabel turunan dan bagaimana cara menganalisanya
2. EDA untuk menemukan suatu pola, mendeteksi anomali, hypothesis testing dan untuk memvalidasi asumsi menggunakan ringkasan statistik dan representasi visual juga digunakan untuk menentukan feature variable yang dapat digunakan oleh machine learning
3. Ragam metode yang berbeda dan tersedia untuk EDA

# Quiz

- Mana dibawah ini yang bukan merupakan teknik dari EDA?
  - A. Univariate visualization
  - B. Bivariate visualization
  - C. Multivariate visualization
  - D. Logistic regression



# Quiz

- Mana dibawah ini yang bukan merupakan teknik dari EDA?
  - A. Univariate visualization
  - B. Bivariate visualization
  - C. Multivariate visualization
  - D. Logistic regression

**Jawaban: D**



## Orbit Future Academy

PT Orbit Ventura Indonesia  
Center of Excellence (Jakarta Selatan)  
Gedung Veteran RI, Lt.15  
Unit Z15-002, Plaza Semanggi  
Jl. Jenderal Sudirman Kav.50, Jakarta  
12930, Indonesia

- 📖 Jakarta Selatan/Pusat
- 📖 Jakarta Barat/BSD
- 📖 Kota Bandung
- 📖 Kab. Bandung
- 📖 Jawa Barat

## Hubungi Kami

Director of Sales & Partnership  
[ira@orbitventura.com](mailto:ira@orbitventura.com)  
**+62 858-9187-7388**

## Social Media



# TERIMA KASIH