

# Statistics

## Homework Guideline



## Estimasi Waktu Pengerjaan



**2 - 3 jam**

## Jumlah Soal



**6 Soal**

## Total Point



**100 poin**

# Teknis Pengerjaan

1. Soal-soal berupa **pertanyaan bisnis** yang dijawab dengan **ilmu statistika**, dan soal-soal dibagi menjadi beberapa bagian berdasarkan tingkat kesulitannya.
  - **Basic** : 4 soal
  - **Advance** : 2 soal
2. Upload hasil pengerjaanmu melalui LMS dengan format nama file sebagai berikut **Nama\_Lengkap\_Batch\_XX** dalam format .html (cara save dalam format .html [disini](#))

# **Soal *Basic***

# Dataset yang digunakan (Soal Basic)

## Home Price Prediction

- **Deskripsi:**

Dataset ini digunakan untuk memprediksi harga rumah. Kita sebagai data scientist diminta tolong, untuk melakukan analisis dengan menggunakan statistika pada data tersebut.

- **Data:**

Setiap baris mewakili rumah, setiap kolom berisi atribut rumah.

- **Link download [disini](#)**



# Soal Pertama (10 poin)

Tunjukkan statistika deskriptif untuk variabel harga rumah (Column Name : *SalePrice*)!

Ceritakan kepada tim bisnis apa makna dari nilai nilai tersebut!

Ekspektasi output :

- ukuran pemusatan (4 poin)
- ukuran penyebaran (4 poin)
- Interpretasi (2 poin)



**5-10 menit**

# Soal Kedua (25 poin)

Tunjukkan statistika deskriptif untuk variabel harga rumah (Column Name : *SalePrice*) setiap tahunnya (Column Name : YrSold)!

Secara deskriptif jelaskan apakah ada perbedaan setiap tahunnya atau tidak?

Ekspektasi output:

- ukuran pemusatan (10 poin)
- ukuran penyebaran (10 poin)
- Interpretasi (5 poin)



# Soal Ketiga (25 poin)

Bagaimana distribusi harga rumah tersebut. Gunakan histogram plot tidak terbatas hanya histogram saja. Jelaskan sebaiknya untuk analisa apakah lebih baik menggunakan mean atau menggunakan median? Beserta alasannya (Column Name : *SalePrice*)

Ceritakan analisa tersebut dengan sudut pandang statistika. Buatlah tim bisnis paham maksud dari grafik histogram tersebut.

Ekspektasi :

- Grafik Histogram (8 poin)
- Analisa (12 poin)
- Interpretasi (5 poin)





# Soal Keempat (30 poin)

**Karena ada pernyataan bahwa setiap tahunnya harga rumah itu pasti naik.** (tahun terjual : YrSold, harga rumah : SalePrice) Silahkan teman-teman buktikan dengan data yang ada menggunakan ilmu statistika? (Tidak terbatas pada pengujian apapun)

Ekspektasi :

- Penentuan hipotesis (5 poin)
- Pengujian yang digunakan (5 poin)
- Hasil pengujian hipotesis (10 poin)
- Kesimpulan dan interpretasi (10 poin)



**20-30 menit**

# **Soal *Advanced***

# AB Test!



# Soal (10 Poin)

Kasus yang akan digunakan adalah **membuka akun deposit untuk nasabah bank**. Pada kesempatan kali ini data scientist diminta tolong untuk menganalisa hasil ab-testing suatu perusahaan bank. Apakah **jumlah campaign** mampu meningkatkan keinginan user untuk **membuka akun deposit** atau tidak?

Data yang digunakan berasal dari folder [bank.zip](#) file yang digunakan adalah **bank-full.csv**

1. Untuk variabel jumlah campaign bisa dicek untuk kolom **campaign**
2. untuk variabel apakah akhirnya user membuka akun atau tidak bisa menggunakan **variabel y**



**50-60 menit**

# Ekspektasi Hasil Analisa

1. Pemahaman konsep A/B Testing **(5 poin)**
  - a. Apa nama experimentnya?
  - b. Definisikan Hipotesis
  - c. Siapa *participant*-nya?
  - d. Variabel yang akan diuji apa?
  - e. Metrics apa yang akan digunakan?
  - f. Berapa sample size dan durasi experiment?
2. Pengujian Hipotesis **(5 poin)**
  - b. Penentuan hipotesis
  - c. Pengujian yang digunakan
  - d. Deteksi asumsi beserta cara penanganannya
  - e. Hasil pengujian hipotesis
  - f. Kesimpulan dan interpretasi



# **Selamat Mengerjakan!**