**LAPORAN EKSPERIMEN**

Pengantar Data Science

**Explorasi Dataset Absenteeism**

Absenteeism\_at\_work.csv

**Disusun Oleh**

William Surya Wijaya | 6182001004

Kae Zarren Luvoma Halim | 6182001005

Ester Oka Nugraheni | 6182001042

**Tanggal**

Sabtu,19 November 2022

**Jurusan Informatika**

**UNIVERSITAS KATOLIK PARAHYANGAN**

1. **Tujuan Eksperimen**

Tujuan dari eksperimen dari data *Absenteeism\_at\_work* adalah untuk mengetahui dan mengenali data berserta hubungan antar atribut untuk mengambil kesimpulan mengenai faktor yang berhubungan dengan ketidakhadiran karyawan dalam perusahaan tersebut.

1. **Deskripsi Dataset, Penyiapan Data, dan Hipotesis**

Atribut data yang ada pada dataset Absenteeism\_at\_work adalah sebagai berikut.

ID

Reason for absence

Month of absence

Seasons

Transportation expense

Distance

Service Time

Age

Workload Avg/Day

Hit Target

Disciplinary failure

Education

Son

Social drinker

Social smoker

Pet

Weight

Height

BMI

Absenteeism in hour

Dataset Absenteeism\_at\_work tidak memilihi nilai NA, null, dan error di setiap atribut yang ada.

Deskripsi dataset dengan visualisasi menggunakan bar-chart untuk mengetahui frekuensi dari nilai tiap atribut.

|  |  |
| --- | --- |
| Reason for absence | Moth of absence |

|  |  |
| --- | --- |
| Seasons | Transportation expense |
| Distance | Service time |
| Age | Workload Avg/Day |

|  |  |
| --- | --- |
| Hit Target | Disciplinary failure |
| Education | Son |
| Social drinker | Social smoker |

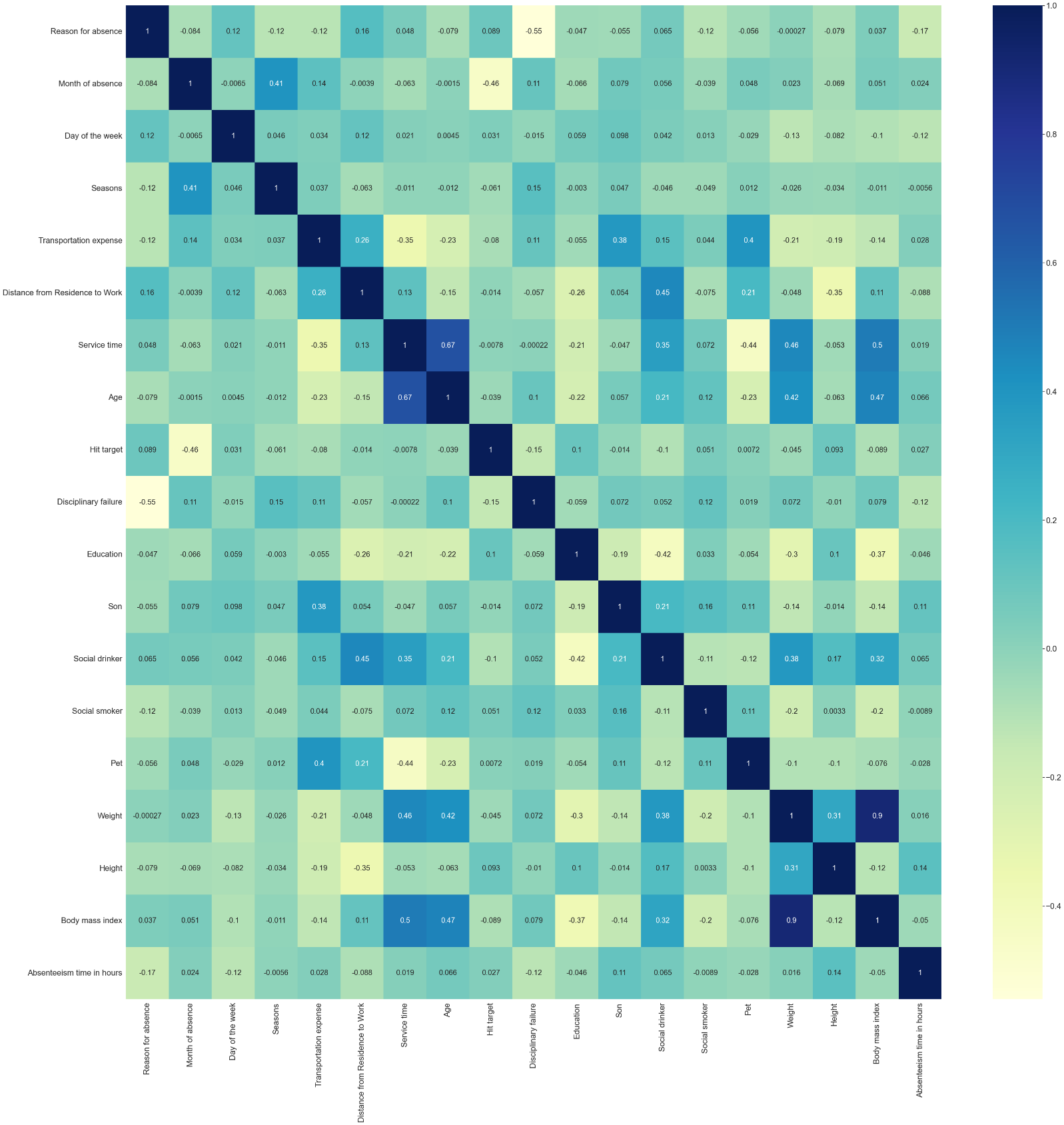
|  |  |
| --- | --- |
| Pet | Weight |
| BMI | Absenteeism in hour |

Hipotesis

1. Variable mana saja yang memiliki pengaruh siginifikan baik hubungan positif dan negatif terhadap atribut target ?
2. Hubungan antar variable mana saja yang memiliki nilai korelasi tinggi ?
3. Penelitian

Untuk mengetahui variable variable yang memiliki pengaruh signifikan baik hubungan positif dan negatif terhadap atribut target adalah salah satunya dengan membuat heatmap.

Pembuatan heatmap dibawah menggunakan function heatmap dari seaborn python.



Untuk nilai korelasi dari tiap atribut dengan atribut target yaitu absenteeism per hour tidak menunjukan nilai korelasi yang tinggi. Variable variable yang memiliki nilai korelasi dengan absenteeism per hour tertinggi adalah sebagai berikut.

|  |  |  |  |
| --- | --- | --- | --- |
| Korelasi Positif | | Korelasi Negatif | |
| Son | 0.11 | Reason for absence | -0.17 |
| Height | 0.14 | Day of the week | -0.12 |
|  |  | Disciplinary failure | -0.12 |

Korelasi Son – Absenteeism per hour (0.11)

Dapat disimpulkan bahwa seorang pegawai yang memiliki anak ( > 0) berpotensi lebih besar untuk absen dengan alasan kepentingan anak.

Korelasi Day of the week – Absenteeism per hour (0.11)

Dapat disimpulkan bahwa seorang pegawai cenderung absen pada hari hari awal dalam sebuah minggu dengan alasan tertentu.

Sementara terdapat nilai korelasi yang tinggi untuk hubungan antar masing masing atribut. Variable variable yang memiliki nilai korelasi antar variable dan dapat dijelaskan hubungannya adalah sebagai berikut.

Korelasi Month of absence – Seasons (0.41)

Urutan bulan (January - December) dan musim (Semi – Dingin) hampir berjalan secara bersama dan teratur, semakin dekat dengan bulan *December* maka musim *Dingin* pun semakin dekat. Dari nilai korelasi tersebut dapat disimpulkan bahwa tingkat absence dalam suatu bulan secara positif dipengaruhi oleh musim.

Korelasi Month of absence – Hit Taget (-0.46)

Hit target dalah target kerja dalam ukuran tertentu yang mewajibkan karyawan untuk mengejar target tersebut dalam jangka waktu tertentu. Dari nilai korelasi tersebut dapat disimpulkan tingkat absence dalam suatu bulan secara negatif dipengaruhi oleh Hit Target.

Korelasi Distance from Residence to Work – Transportation Expense (0.26)

Jarak dari tempat tinggal menuju tempat kerja mempengaruhi biaya transportasi dikarenakan mempengaruhi biaya bahan bakar untuk pengguna kendaraan pribadi dan biaya jasa transportasi untuk pengguna kendaraan umum. Dari nilai korelasi tersebut dapat disimpulkan biaya transportasi secara positif dipengaruhi oleh jarak dari tempat tinggal ke tempat kerja.

Korelasi Age – Service Time (0.67)

Korelasi antar age dan service time dapat menunjukan lamanya seseorang bekerja dalam perusahaan tersebut. Dari nilai korelasi tersebut dapat disimpulkan bahwa service time secara positif dipengaruhi oleh umur.

Lampiran

Visualisasi Data (chart & heatmap)

[*https://github.com/shrk-sh-ioai/pds-banking-dt/tree/main/tugas-absence/assets*](https://github.com/shrk-sh-ioai/pds-banking-dt/tree/main/tugas-absence/assets)