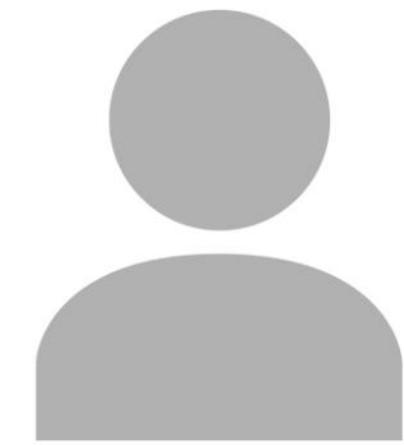


# Improving Employee Retention by Predicting Employee Attrition Using Machine Learning

Supported by:  
**Rakamin Academy**  
Career Acceleration School  
[www.rakamin.com](http://www.rakamin.com)



**Created by:**  
**Candraditya Dwaya Putra**  
Email : [candradityaputra1@gmail.com](mailto:candradityaputra1@gmail.com)  
**linkedIn Profile :**  
<https://www.linkedin.com/in/candraditya-dwaya-putra-77719a66>

“Have more than 11 years professional career in GIS and Management geodatabase. Good knowledge WebGIS Development, with strong experience Geospatial data in Forestry, environmental and regional planning. Skilled at data collection and analysis that elicits accurate and valuable information utilizing technical principles and theories. Technical proficiencies include SQL, Python, Machine Learning, MS Office, ESRI GIS, QGIS, HTML, CSS/Bootstrap, Javascript, geomorphology and Geospatial Software.”

# Overview

“Sumber daya manusia (SDM) adalah aset utama yang perlu dikelola dengan baik oleh perusahaan agar tujuan bisnis dapat tercapai dengan efektif dan efisien. Pada kesempatan kali ini, kita akan menghadapi sebuah permasalahan tentang sumber daya manusia yang ada di perusahaan. Fokus kita adalah untuk mengetahui bagaimana cara menjaga karyawan agar tetap bertahan di perusahaan yang ada saat ini yang dapat mengakibatkan bengkaknya biaya untuk rekrutmen karyawan serta pelatihan untuk mereka yang baru masuk. Dengan mengetahui faktor utama yang menyebabkan karyawan tidak merasa, perusahaan dapat segera menanggulanginya dengan membuat program-program yang relevan dengan permasalahan karyawan.”

# Dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 287 entries, 0 to 286
Data columns (total 25 columns):
 #   Column           Non-Null Count  Dtype  
 ---  -- 
 0   Username        287 non-null    object  
 1   EnterpriseID   287 non-null    int64  
 2   StatusPernikahan 287 non-null    object  
 3   JenisKelamin    287 non-null    object  
 4   StatusKepegawaian 287 non-null    object  
 5   Pekerjaan       287 non-null    object  
 6   JenjangKarir    287 non-null    object  
 7   PerformancePegawai 287 non-null    object  
 8   AsalDaerah      287 non-null    object  
 9   HiringPlatform   287 non-null    object  
 10  SkorSurveyEngagement 287 non-null    int64  
 11  SkorKepuasanPegawai 282 non-null    float64 
 12  JumlahKeikutsertaanProjek 284 non-null    float64 
 13  JumlahKeterlambatanSebulanTerakhir 286 non-null    float64 
 14  JumlahKetidakhadiran   281 non-null    float64 
 15  NomorHP          287 non-null    object  
 16  Email            287 non-null    object  
 17  TingkatPendidikan 287 non-null    object  
 18  PernahBekerja    287 non-null    object 
```

## DESCRIPTION

- Dataset terkait status karyawan termasuk tanggal masuk dan keluar karyawan

## SHAPE

- 287 Row and 25 Columns

## DTYPES

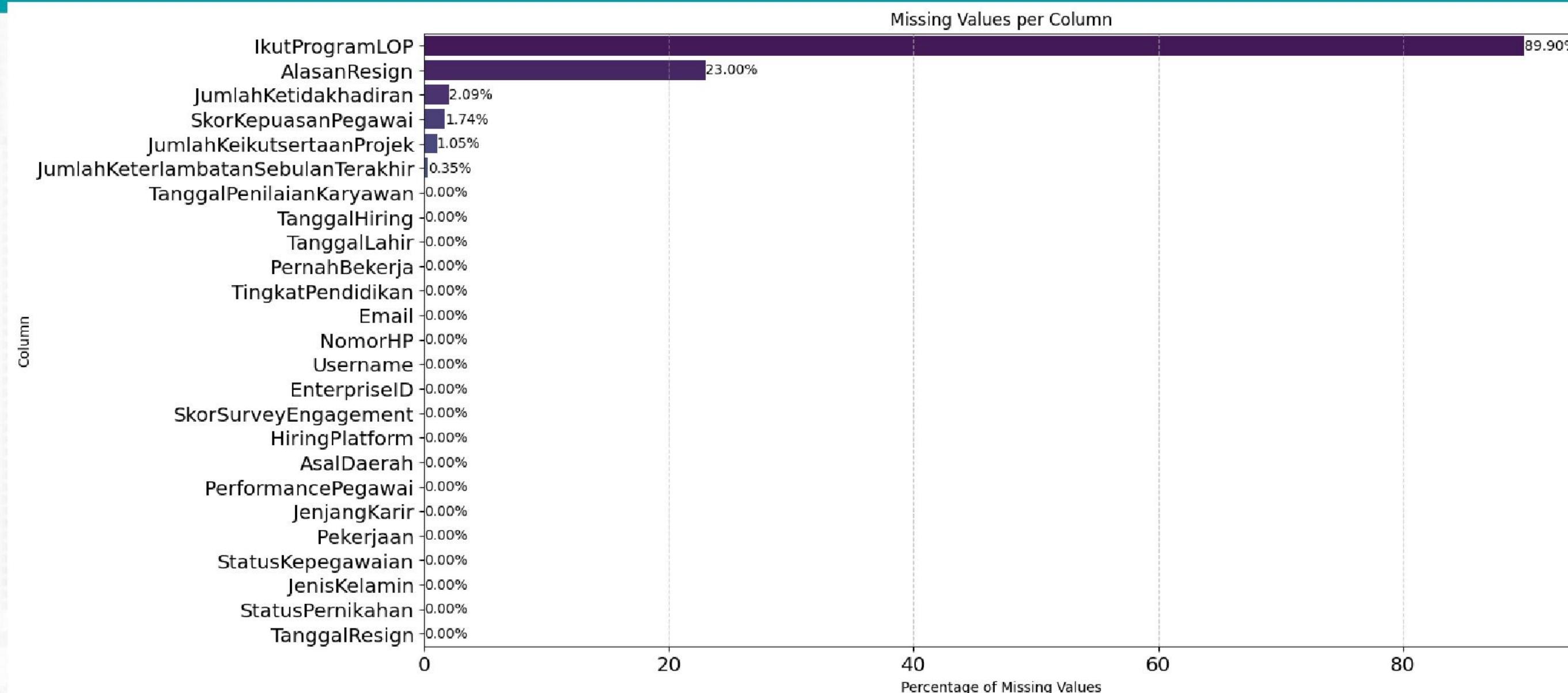
- float64(5), int64(2), object(18)

## MISSING VALUES

- 6 features yang memiliki missing value, yaitu SkorKepuasanPegawai, JumlahKeikutsertaanProjek, JumlahKeterlambatanSebulanTerakhir, JumlahKetidakhadiran, IkutProgramLOP dan AlasanResign

Untuk selengkapnya, dapat melihat jupyter notebook disini

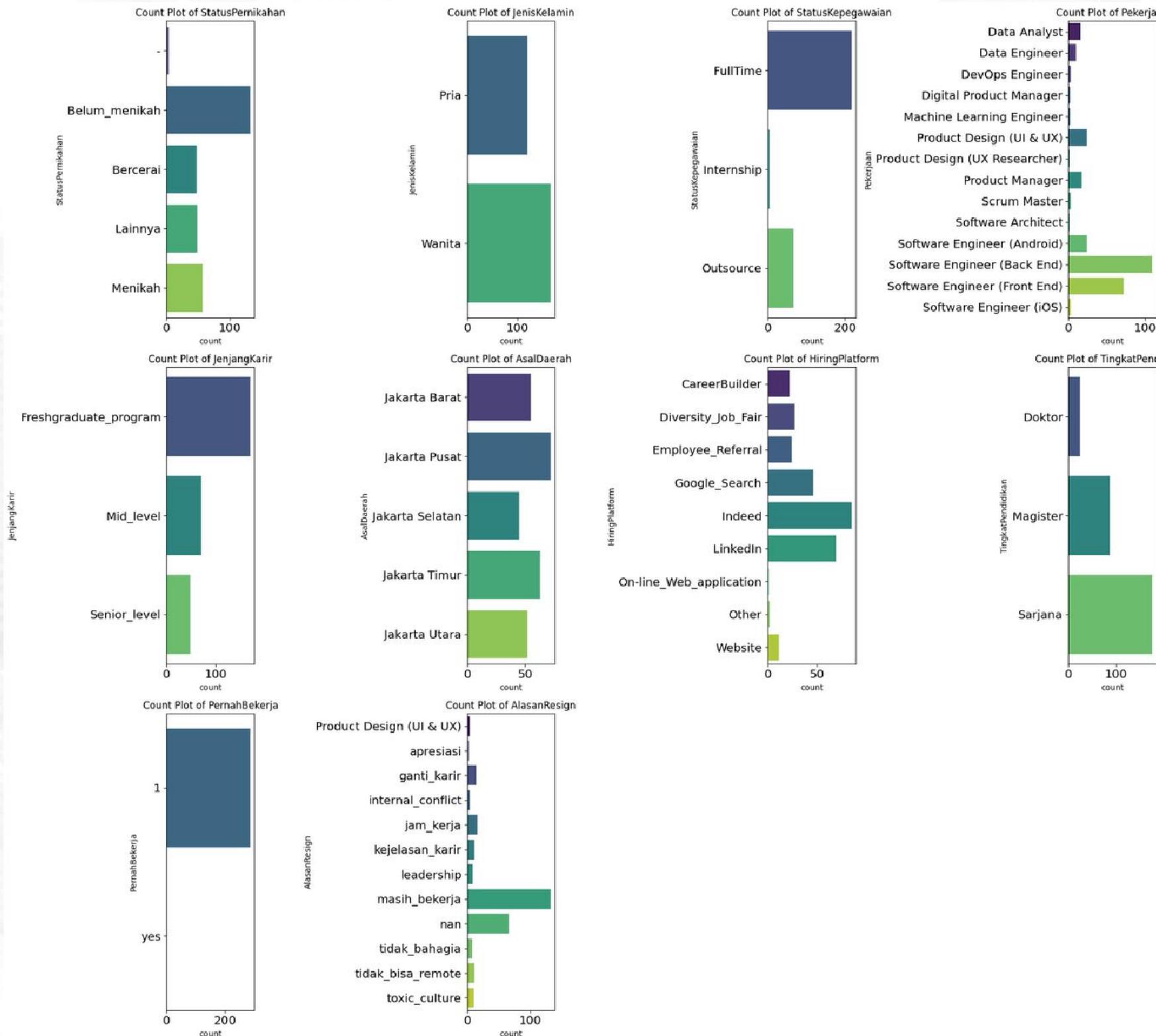
# Data Preprocessing



## • Handling Missing Value

feature yang dihandling SkorKepuasanPegawai, JumlahKeikutsertaanProjek, JumlahKeterlambatanSebulanTerakhir, JumlahKetidakhadiran, IkutProgramLOP dan AlasanResign

# Data Preprocessing



## Handling Inconsistent data

Jika dilihat pada diagram diatas, untuk berbagai kolom terdapat data yang harus dirapikan, seperti:

- for marital status column:  
memindahkan nilai data '-' yang digabungkan ke 'other'
- Column Ever worked:  
menggabungkan yang memiliki nilai 1 = 'yes'
- column for reason of resignation:  
menggabungkan product design (UI & UX) into 'other'  
selanjutnya akan dibuat kolom baru terkait status berfungsi atau tidak

## Feature Engineering

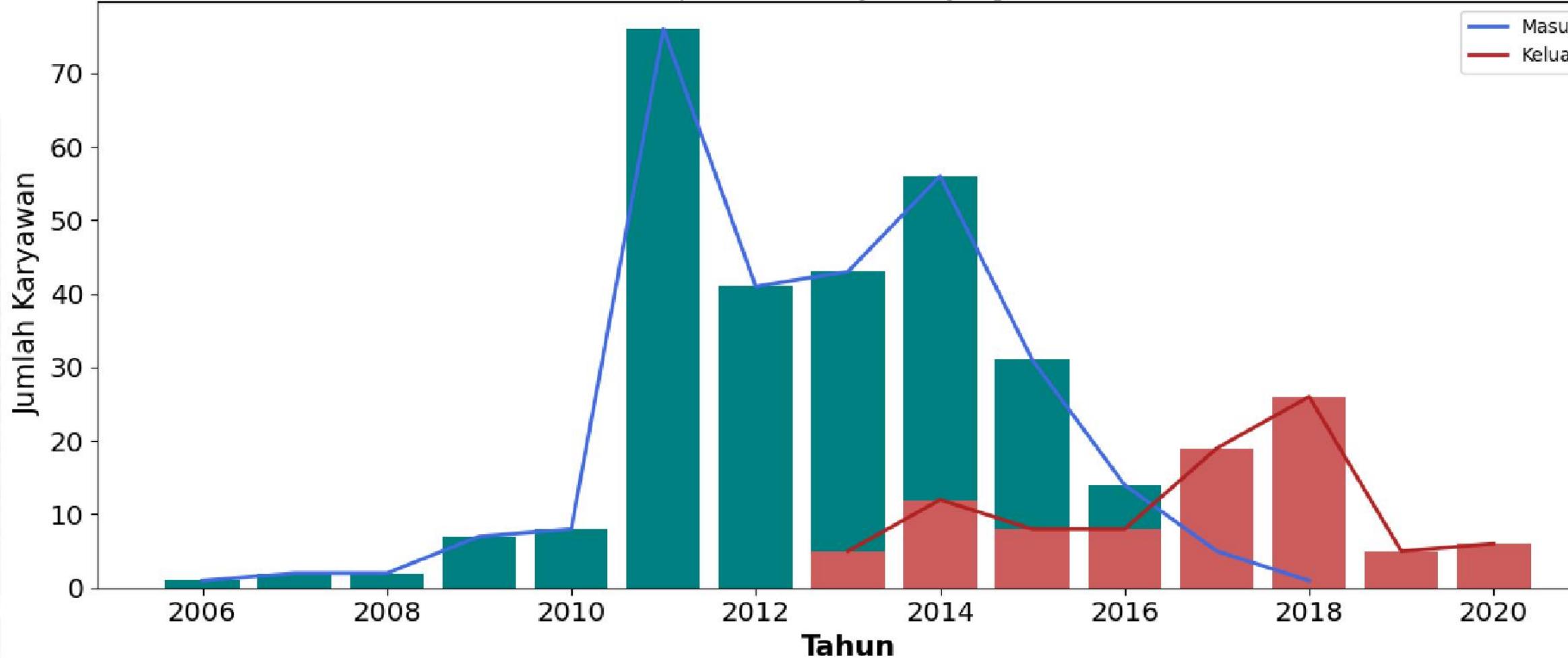
- Status Kerja  
memindahkan nilai data '-' yang digabungkan ke 'other'
- Kolom Hiring  
menggabungkan yang memiliki nilai 1 = 'yes'
- Tanggal lahir:
- Tanggal penilaian Karyawan

Untuk selengkapnya, dapat melihat jupyter notebook disini  
<https://drive.google.com/file/d/1GpvuX54CgGgfLXF40Qryp3QzgYZBK89P/view?usp=sharing>

# Annual Report on Employee Number Changes

## Tren Perubahan Jumlah Karyawan Tiap Tahun

Terjadi low point saat tahun 2018 dimana hampir 20 lebih karyawan resign serta penurunan karyawan yang masuk

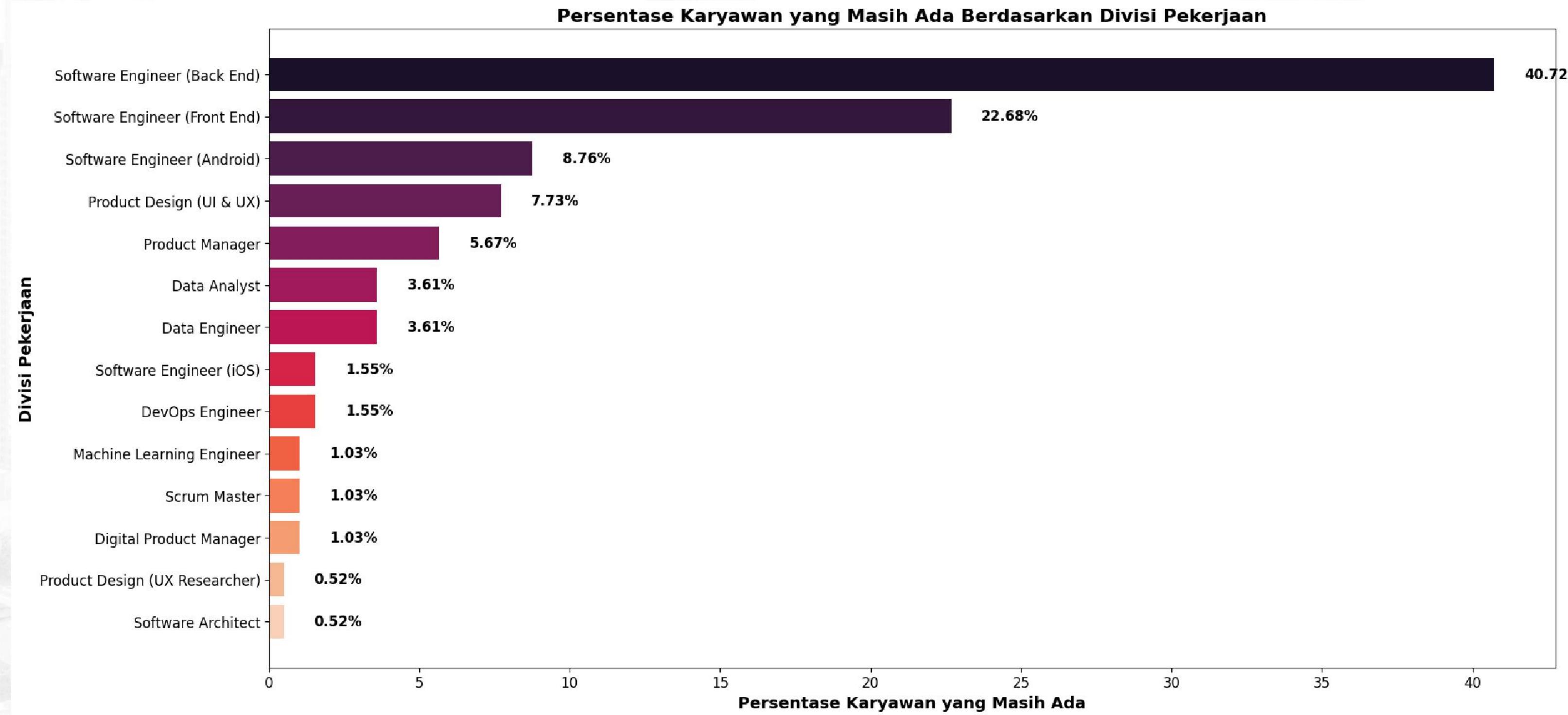


- Berdasarkan data jumlah pegawai yang masuk dan keluar setiap tahunnya, dapat diambil beberapa wawasan mengenai kondisi dunia kerja pegawai. Berikut wawasan yang bisa diperoleh:
- Jumlah pegawai masuk mengalami peningkatan yang signifikan dari tahun 2006 hingga tahun 2011. Namun setelah tahun 2011, tren pertumbuhan jumlah pegawai masuk cenderung menurun secara perlahan.
- Jumlah pegawai yang keluar juga mengalami peningkatan dari tahun ke tahun. Tren peningkatan jumlah pegawai yang keluar dapat menjadi perhatian karena dapat mengindikasikan faktor-faktor yang mempengaruhi keputusan pegawai untuk mengundurkan diri.
- Perubahan jumlah pegawai (selisih jumlah incoming Employees dan Outgoing Employees) berfluktuasi dari tahun ke tahun. Ada tahun-tahun dimana perubahan jumlah pegawai bernilai negatif, yang menunjukkan adanya penurunan jumlah pegawai secara keseluruhan.
- Berdasarkan pemaparan di atas, kondisi angkatan kerja pegawai mungkin mengkhawatirkan karena adanya tren penurunan pertumbuhan jumlah pegawai yang masuk dan peningkatan jumlah pegawai yang keluar.

Untuk selengkapnya, dapat melihat jupyter notebook disini

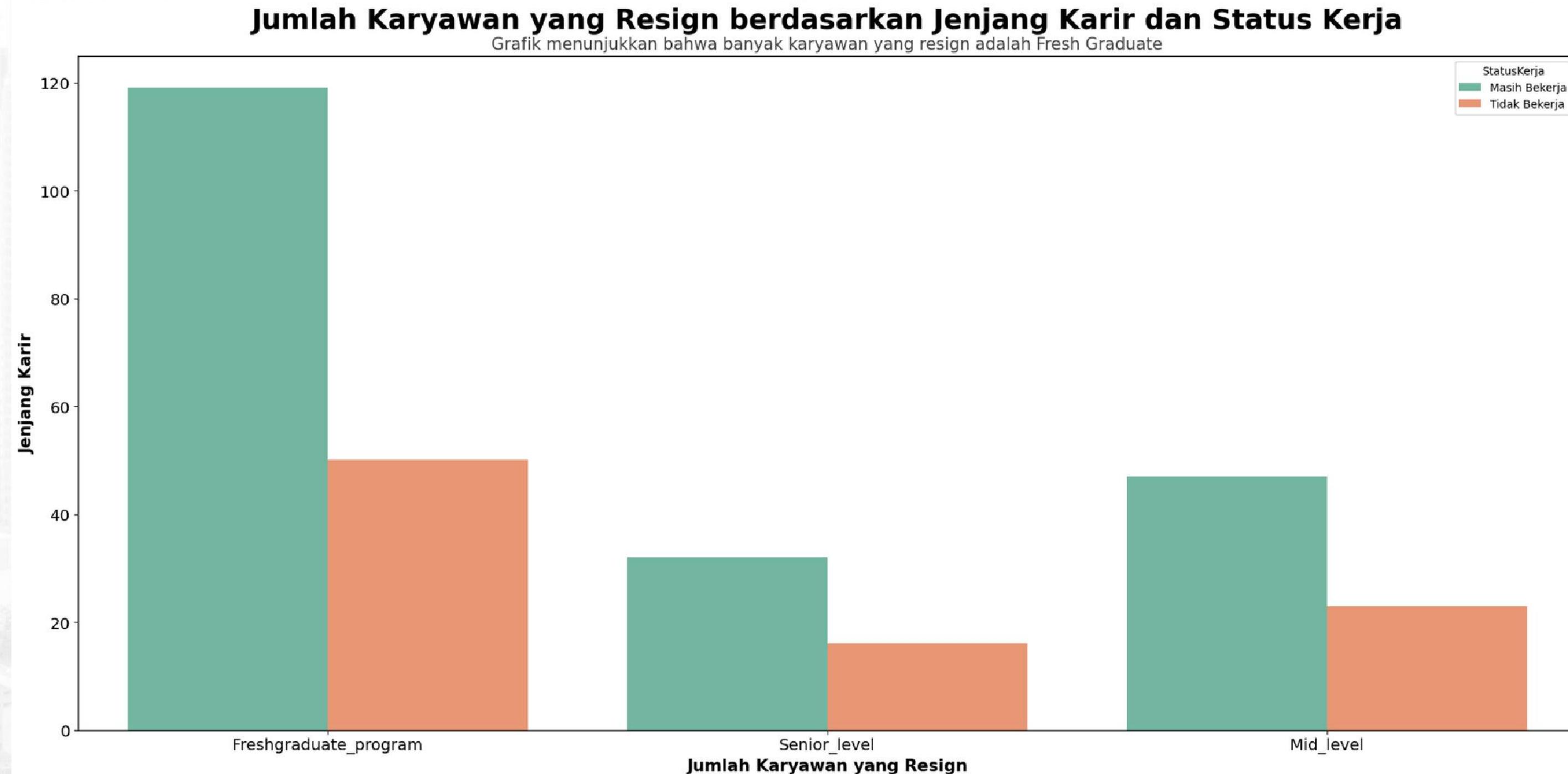
<https://drive.google.com/file/d/1GpvuX54CgGgfLXF40Qryp3QzgYZBK89P/view?usp=sharing>

# Annual Report on Employee Number Changes



- Software Engineering karyawan paling dominan di perusahaan backend, frontend dan android

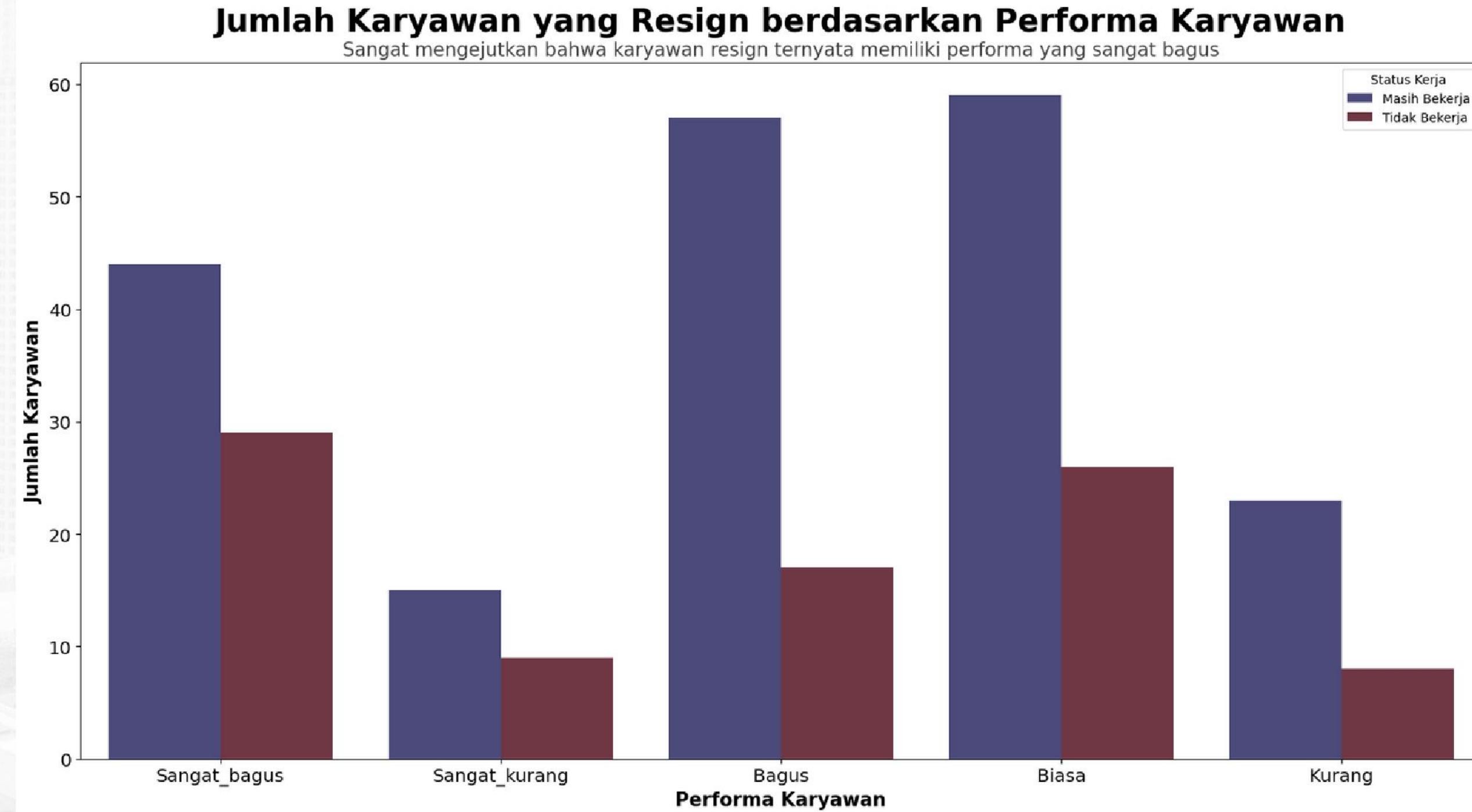
# Resign Reason Analysis For Employee Attrition Management Strategy



- Freshgraduate\_program menunjukkan jumlah paling banyak resign pada perusahaan

Untuk selengkapnya, dapat melihat jupyter notebook disini  
<https://drive.google.com/file/d/1GpvuX54CgGgfLXF40Qryp3QzgYZBK89P/view?usp=sharing>

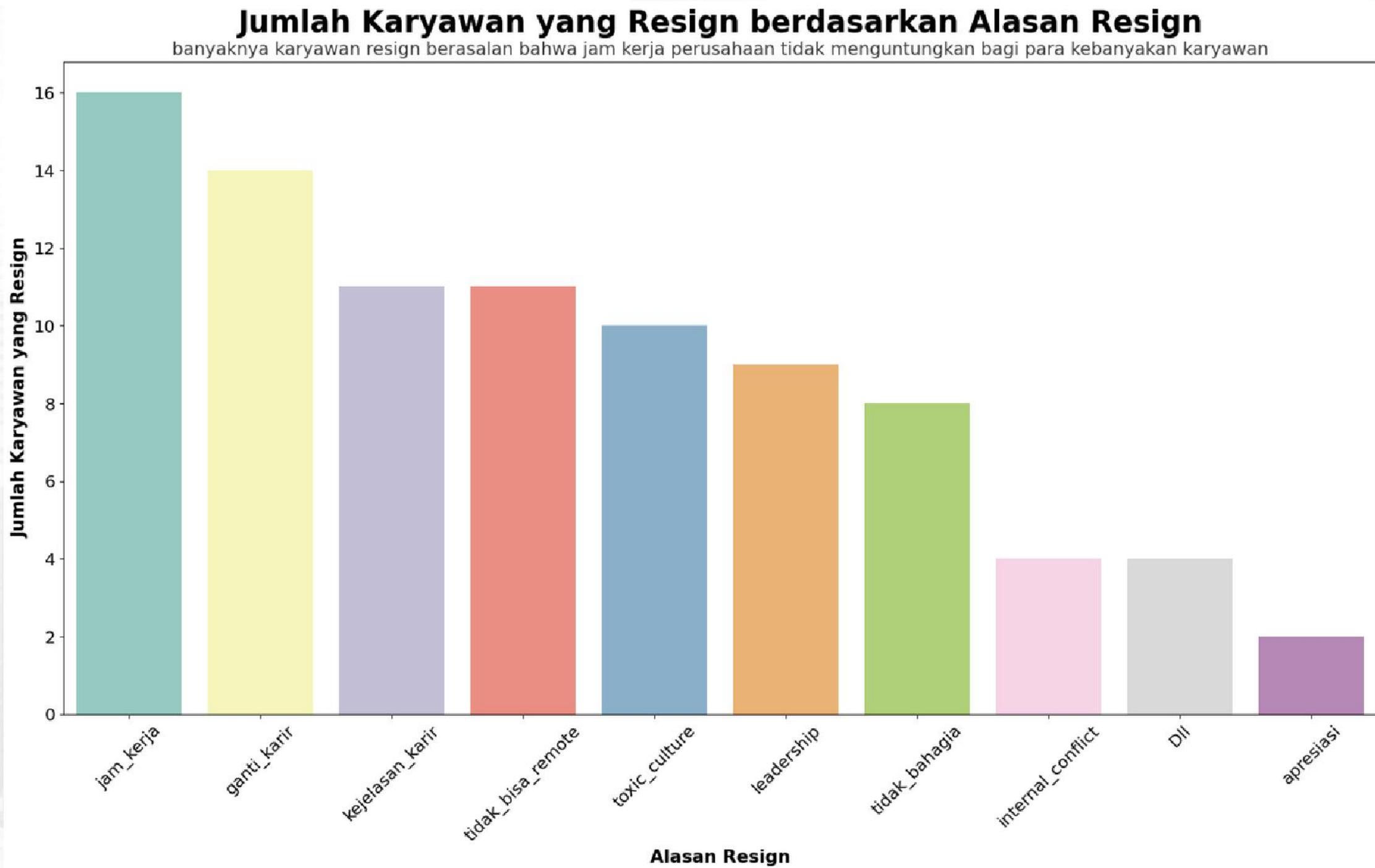
# Resign Reason Analysis For Employee Attrition Management Strategy



- Performa sangat\_bagus menunjukkan jumlah paling banyak resign pada perusahaan diikuti oleh performa Biasa lalu Bagus

Untuk selengkapnya, dapat melihat jupyter notebook disini  
<https://drive.google.com/file/d/1GpvuX54CgGgfLXF40Qryp3QzgYZBK89P/view?usp=sharing>

# Resign Reason Analysis For Employee Attrition Management Strategy



- PAlasan mereka resign di dominasi oleh alasan Jam\_kerja lalu diikuti ganti\_karir dan menginginkan pekerjaan remote yang ditunjukkan pada feature tidak\_bisa\_remote

Untuk selengkapnya, dapat melihat jupyter notebook disini

<https://drive.google.com/file/d/1GpvuX54CgGgfLXF40Qryp3QzgYZBK89P/view?usp=sharing>

# Build An Automated Resignation Behavior Prediction Using Machine Learning

Mean Precision: 0.97 ( $\pm 0.05$ )

Mean Recall: 0.96 ( $\pm 0.07$ )

Mean ROC-AUC: 0.97 ( $\pm 0.03$ )

- Hasil cross validation menunjukkan bahwa model tersebut menunjukkan skor presisi dan perolehan yang kuat, yang menunjukkan kemampuannya dalam mengklasifikasikan kasus positif dengan benar sambil meminimalkan positif palsu dan negatif palsu. Selain itu, skor ROC-AUC yang tinggi mencerminkan kemampuan diskriminasi model yang sangat baik secara keseluruhan.

# Build An Automated Resignation Behavior Prediction Using Machine Learning

Best Hyperparameters:

```
{'learning_rate': 0.01, 'max_depth': 4, 'min_samples_leaf': 5, 'min_samples_split': 3, 'n_estimators': 50, 'subsample': 0.8}  
Best Mean ROC-AUC: 1.00
```

- Hasil Hyperparameter Set menunjukkan bahwa model tersebut menunjukkan kekuatan prediksi yang luar biasa, dengan skor ROC-AUC mencapai nilai tertinggi sebesar 1,00. Pilihan hyperparameter dan skor ROC-AUC yang dihasilkan menggarisbawahi kemampuan model yang kuat dalam membedakan kasus positif dan negatif dengan diskriminasi sempurna. Singkatnya, model ini, yang telah disesuaikan dengan hyperparameter yang teridentifikasi, unggul dalam performa klasifikasinya dan sangat cocok untuk tugas yang memerlukan akurasi prediksi tinggi.
- Untuk selengkapnya, dapat melihat jupyter notebook disini <https://drive.google.com/file/d/1GpvuX54CgGgfLXF40Qryp3QzgYZBK89P/view?usp=sharing>

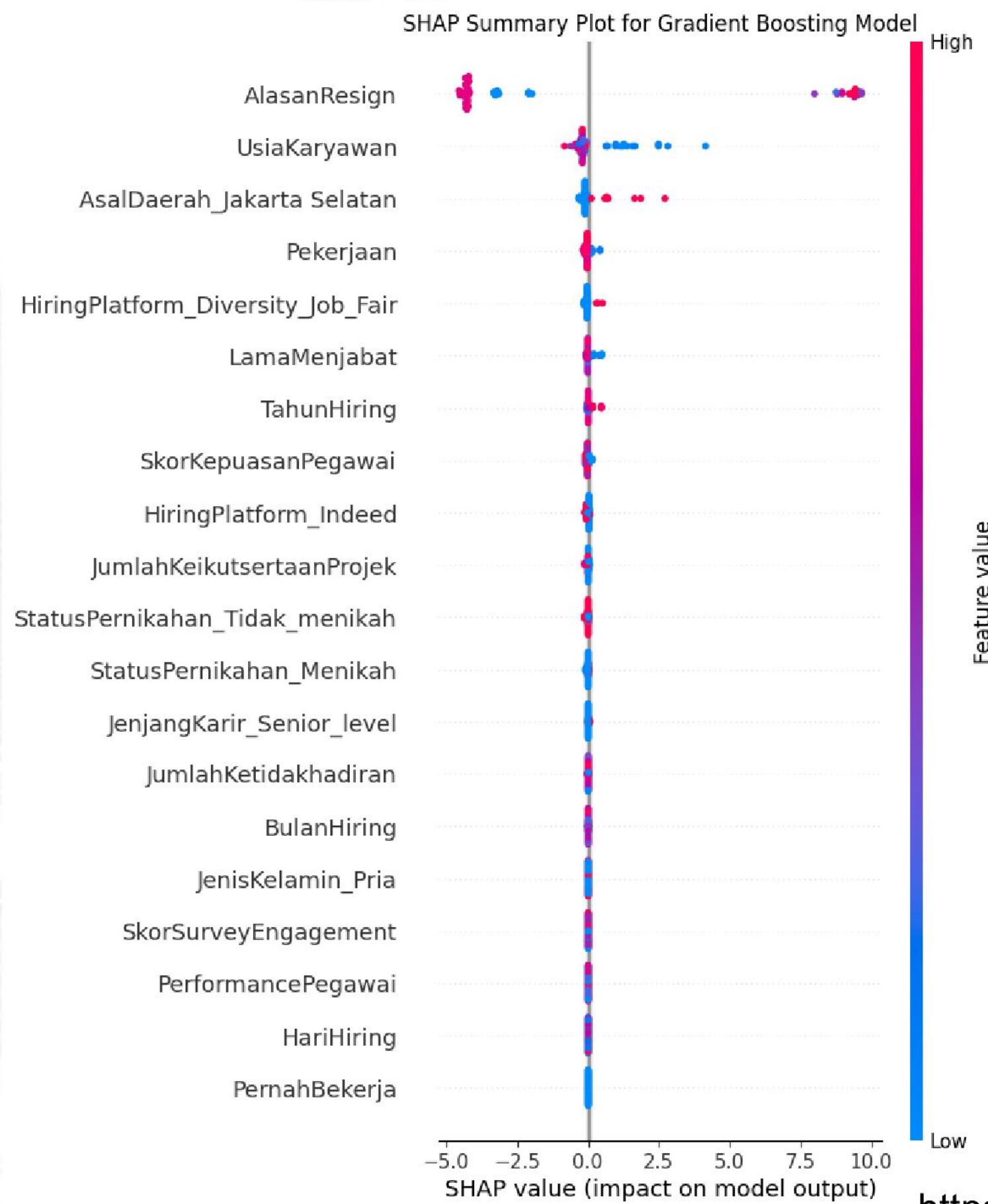
# Build An Automated Resignation Behavior Prediction Using Machine Learning

Best Hyperparameters:

```
{'learning_rate': 0.01, 'max_depth': 4, 'min_samples_leaf': 5, 'min_samples_split': 3, 'n_estimators': 50, 'subsample': 0.8}  
Best Mean ROC-AUC: 1.00
```

- Hasil Hyperparameter Set menunjukkan bahwa model tersebut menunjukkan kekuatan prediksi yang luar biasa, dengan skor ROC-AUC mencapai nilai tertinggi sebesar 1,00. Pilihan hyperparameter dan skor ROC-AUC yang dihasilkan menggarisbawahi kemampuan model yang kuat dalam membedakan kasus positif dan negatif dengan diskriminasi sempurna. Singkatnya, model ini, yang telah disesuaikan dengan hyperparameter yang teridentifikasi, unggul dalam performa klasifikasinya dan sangat cocok untuk tugas yang memerlukan akurasi prediksi tinggi.
- Untuk selengkapnya, dapat melihat jupyter notebook disini <https://drive.google.com/file/d/1GpvuX54CgGgfLXF40Qryp3QzgYZBK89P/view?usp=sharing>

# Presenting Machine Learning Products To The Business Users



- Berdasarkan Gradieny Boostin Model, AlasanResign dan UsiaKaryawan menjadi penyebab karyawan keluar dari perusahaan

Untuk selengkapnya, dapat melihat jupyter notebook disini

<https://drive.google.com/file/d/1GpvuX54CgGgfLXF40Qryp3QzgYZBK89P/view?usp=sharing>