# 

# **PT. PAHAMI CIPTA EDUKASI (PAHAMIFY)**

**TECHNICAL TEST : DATA ANALYST**

**[Dokumen Rahasia]**

***Hasil Test Anda hanya digunakan untuk kebutuhaan proses recruitment.***

***Untuk menjawab tidak perlu request to edit, cukup make a copy dari file ini.***

***Saat mengirimkan file mohon untuk filenya tidak di protek, supaya bisa di akses oleh reviewer.***

**Nama Pelamar : Jafar Shodiq**

**No. Handphone : +6281382378203**

**Email : j.shodiq2312@gmail.com**

# **[Soal]**

Data yang kami gunakan untuk merupakan data yang sudah di staging dan dalam bentuk dummy (bukan data sebenarnya), sehingga hasil tes ini tidak bisa digunakan untuk kepentingan perusahaan. Test ini hanya untuk mengecek kemampuan teknikal kandidat.

1. Jelaskan dan sebutkan, matriks apa saja yang harus diperhatikan oleh startup yang bergerak di bidang EdTech untuk membuat growth perusahaan? Tentukan yang menjadi success matrik dan tracking metrics, Dan bagaimana cara mengkuantifikasikan matrik tersebut?
2. Bagaimana cara menentukan marketing activity (online & offline) itu berhasil dan efektif, apa yang menjadi KPI nya? Bagaimana cara mengukurnya? Dan tools apa yang dapat digunakan untuk mengukur hal tersebut?
3. Silahkan Query (dapat menggunakan SQL, MySQL, PostgreSQL, Google Spreadsheet Query, dll) data-data ini: (**Data ada di Notes**)
   1. Buatlah Query untuk mendapatkan **Data User yang Statusnya Premium** dalam 1 table dan berikan screenshot hasilnya
   2. Buatlah Query untuk mendapatkan **Data User yang Mendaftar Try Out UTBK dengan Akses Level Free, Full, dan Video** dalam 1 table dan berikan screenshot hasilnya
   3. Buatlah Query untuk mendapatkan **Data User yang Membeli Subscription Packages (Einstein, Marie Curie, tesla, & Newton)** dan Harga Bersih yang Didapatkan dalam 1 table dan berikan screenshot hasilnya
   4. Buatlah query untuk mendapatkan **Total Revenue untuk Paket Subscription** sesuai dengan **tingkat Kelasnya (kelas X, XI, & XII)** dalam 1 table dan berikan screenshot hasilnya
   5. Buatlah Query untuk mendapatkan **Nama Paket Subscription Beserta Jumlah Murid Kelas X, XI, XII, dan seluruh murid yang membeli sesuai paketnya (Einstein, Marie Curie, Tesla, & Newton). Total Revenue dan rataannya berdasarkan penjualan pada kelas X, XI, XII, dan keseluruhan** dalam 1 Table dan sertakan screenshot table beserta isi tablenya
   6. *Explore*-lah sebanyak-banyaknya **Query dan berikan screenshot** yang bisa kalian dapatkan dari data yang telah diberikan dan gunakanlah teknik-teknik serta jelaskan cara penggunaan teknik tersebut (**Contoh teknik: Pivot Table**)
   7. (**Bonus**)Dari 3 Proses ETL (Extraction, Transform, & Load), Jelaskan dan berikanlah **step by step** untuk melakukan **Load** dalam proses ETL dengan kasus data yang telah diberikan diatas.
4. Buatlah proyeksi growth pahamify dari data yang tersedia diatas, mengacu pada matriks-matriks yang kamu sebutkan pada no 1, kamu bisa menggunakan konsep data science apapun, dengan software analisis statistik apapun, dan buatlah data visualisasinya (dapat menggunakan google data studio, tableu ataupun power BI) !
5. Tuliskan insight apa saja yang kamu dapat dari data diatas dan berikan action plan yang harus dikerjakan oleh internal tim pahamify, dan jelaskan mana yang harus dijadikan prioritas dari insight tersebut.
6. Buatlah Competitor Analysis berdasarkan social media maupun google playstore, dan tuliskan insight dan rekomendasinya, buatlah dalam bentuk presentasi yang menarik dalam slide presentasi! Kamu boleh melakukan Competitor analysis dengan menggunakan startup Edtech sebagai objek maupun startup lainnya

# **[Notes]**

1. Link untuk data bisa dibuka di link berikut: <https://drive.google.com/drive/folders/12WmshMwBlvQaTT1SO5xFNGtsIsAPWdG6?usp=sharing>
2. Contoh Bentuk Table Untuk Soal **nomor 3.e**:



1. Jika menemukan data yang berisi **Default** pada **user\_profile,** itu berarti user tidak menyelesaikan atau tidak mengisi data **profile** miliknya
2. Jika pada satu kolom hanya terdapat angka **1** dan **0** itu berarti menunjukkan bahwa pada kolom tersebut terdapat sebuah variabel boolean berupa **true (1)** & **false (0)**
3. **Top Up** berarti user melakukan Isi ulang terhadap saldo sedangkan **Charge** berarti User melakukan pembelian atau pembayaran terhadap suatu hal
4. Pada Transaction Status:
   1. **Init** → Berarti User melakukan inisialisasi pertama kali terhadap suatu barang
   2. **deny** → Berarti Pembelian suatu User ditolak oleh provider
   3. **Expire** → Berarti Pembelian telah melebihi batas pembayaran yang telah ditentukan oleh sistem
   4. **Settlement** → Pembayaran telah berhasil dan diterima oleh Provider
5. Pada Status Level Participant Terdapat 3 buah parameter:
   1. **Free** → User yang mengikuti tes secara gratis dengan syarat yang telah disediakan untuk menaikan follower pada Sosial Media
   2. **Video** → User yang mengikuti tes namun membayar secara khusus untuk mendapatkan Video berupa pembahasan tes
   3. **Full** → User yang mengikuti tes secara berbayar sehingga mendapatkan soal dan video pembahasan yang khusus

# **[Jawab]**

1. Jawaban 1

|  |
| --- |
| Success metrics: Average Revenue Per Account (ARPA), Net Promoter Score (NPS)  Tracking metrics: Churn Rate, Customer Satisfaction Score (CSAT), Monthly Recurring Avenue (MRR)  ARPA = (MRR / # customer)  NPS = %promoters - %detractors  Churn Rate = (# customer churn di periode tertentu / # customer churn di awal periode tertentu)  CSAT(%) = (# respon positif / # total respon) x 100  MRR = sum(revenue berulang) |

1. Jawaban 2

|  |
| --- |
|  |

1. Jawaban 3

Untuk bagian ini, karena data juga disediakan berupa file CSV, maka asumsi saya adalah saya boleh memakai salah satu format file (CSV atau SQL). Dalam kasus saya, saya lebih nyaman untuk melakukan analisa data dengan menggunakan tools python seperti pandas. Hal ini bukan berarti saya tidak familiar sama sekali dengan SQL, hanya masalah preferensi dan melihat efisiensi waktu pengerjaan yang terbilang tidak banyak. Jadi dimohon untuk pengertiannya.

Kolom jawaban `Query` akan saya isi dengan baris kode yang saya lakukan di jupyter notebook dengan bahasa pemrograman python.

* 1. Bagian A

|  |
| --- |
|  |
| df = pd.read\_csv('./CSV/canva\_transaction/user\_status.csv', sep=';')  df[df['UserPremiumStatus']==1] |

* 1. Bagian B

|  |
| --- |
| FULL    FREE    PENDING    VIDEO |
| df = pd.read\_csv('./CSV/canva\_test\_participant/participant.csv', sep=';')  df['StatusLevelParticipant'].value\_counts()  df[df['StatusLevelParticipant']=='FULL']  df[df['StatusLevelParticipant']=='FREE']  df[df['StatusLevelParticipant']=='PENDING']  df[df['StatusLevelParticipant']=='VIDEO'] |

* 1. Bagian C

|  |
| --- |
|  |
| df[df['ActionName'].str.contains('Einstein')]  eins = df['TotalPrice'][df['ActionName'].str.contains('Einstein')].sum()  marie = df['TotalPrice'][df['ActionName'].str.contains('Marie Curie')].sum()  tesla = df['TotalPrice'][df['ActionName'].str.contains('Tesla')].sum()  newt = df['TotalPrice'][df['ActionName'].str.contains('Newton')].sum()  print("Einstein total price: {}".format(eins))  print("Marie Curie total price: {}".format(marie))  print("Tesla total price: {}".format(tesla))  print("Newton total price: {}".format(newt)) |

* 1. Bagian D

|  |
| --- |
| Informasi tidak memadai. Selain itu, file sql tidak menunjukkan foreign key sehingga cukup sulit untuk saya menentukan tabel mana yang berkorelasi satu sama lain. |
| Query |

* 1. Bagian E

|  |
| --- |
| Informasi tidak memadai. Selain itu, file sql tidak menunjukkan foreign key sehingga cukup sulit untuk saya menentukan tabel mana yang berkorelasi satu sama lain. |
| Query |

* 1. Bagian F

|  |
| --- |
| Gender    District |
| df = pd.read\_csv('./CSV/canva\_users/user\_profile.csv', sep=';')  df ['Gender'].value\_counts()  df['District'].value\_counts().head()  len(df[df['SchoolYear']=='XI IPA']) |

* 1. Bagian G

|  |
| --- |
| Penjelasan Load dalam ETL |
| Screenshot Step-by-Step |
| Penjelasan Step-By-Step |

1. Jawaban 4

|  |
| --- |
|  |

1. Jawaban 5

|  |
| --- |
| Level participant    Update terakhir participant (pendaftaran(?)) paling banyak di tahun 2019, bulan desember, dan di pertengahan bulan terutama tanggal 19.    Tidak banyak transaksi topup karena topup biasa dilakukan sesekali tapi langsung banyak ammountnya.    Transaction status |

1. Jawaban 6

|  |
| --- |
|  |