

Capstone Project MyEduSolve



Presented by Group 3
Data Analyst Pathway



OUR GROUP MEMBER



**Adventus Totti
Mariano Simbolon**

Universitas Mikroskil



**Claryta Putri Dedyana
Wati**

Universitas Airlangga



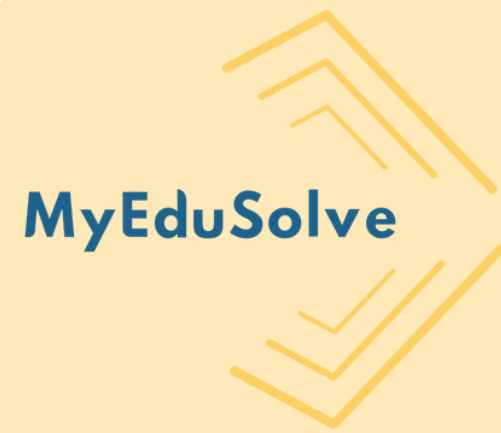
**Doni Ismail
Awang Santoso**

UPN Veteran Jawa Timur



**Ray Gerraldo S
Tambunan**

Universitas Negeri Semarang



MyEduSolve is an educational technology company whose mission is to elevate Indonesia's workforce through certification and online learning.

To evaluate the quality of digital learning services provided by MyEduSolve, a survey was conducted on several programs :

- Data Science Track
- Career Ready Track
- Microsoft Office Specialist Expert
- Adobe Certified Professional in Video Design
- Adobe Certified Professional in Visual Design

And from this survey, MES team can find out the classification personality of each user that fill survey



What personality type will be classified?





Introvert

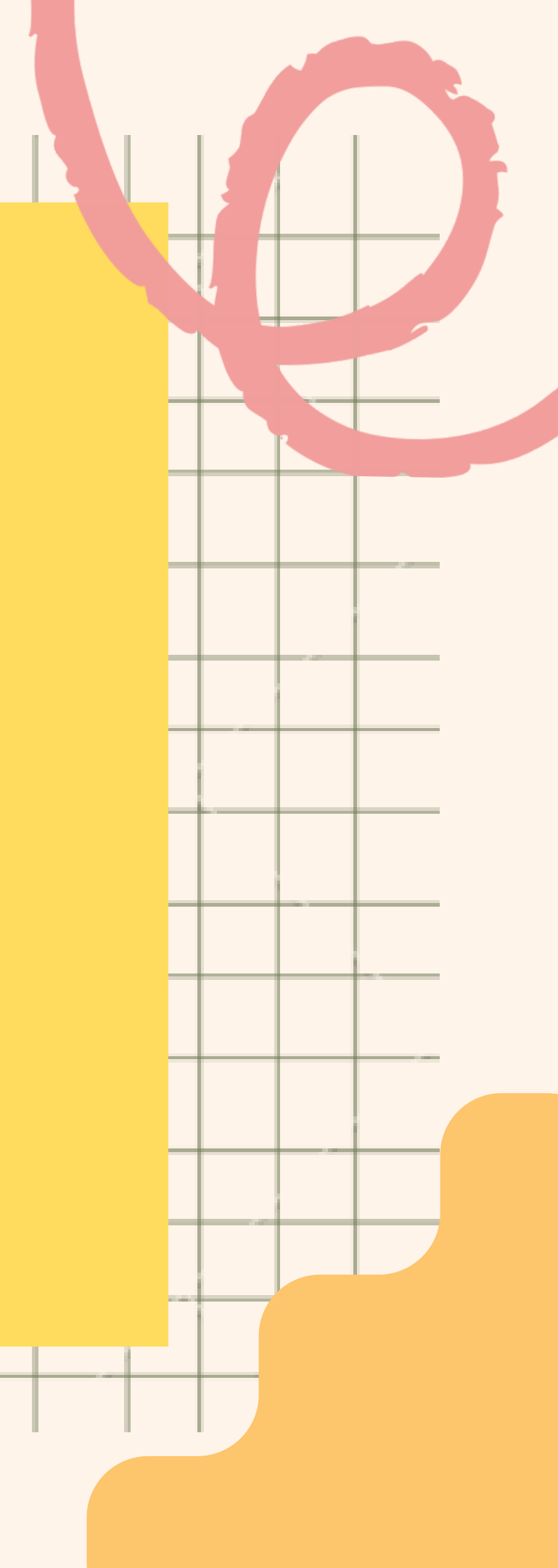
Introverts are personalities type that tend to focus more than within themselves, thoughts that require more time in their own space.





Extrovert


Extroverts are personalities that are more likely to enjoy active free space by talking to the people around them and mostly Extrovert people can be multitasking.





Career Ready

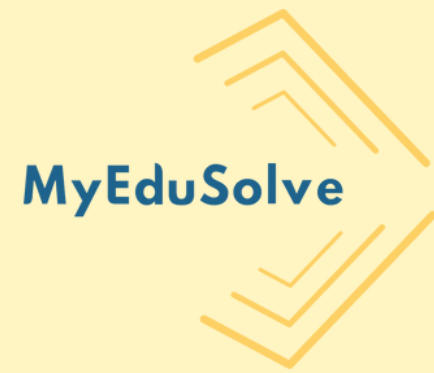
Career ready are personalities trait of someone who joins the program because they want to deepen skills that match their passion.





Career Not Ready

Career not ready are personalities type where they still confused about what their goal is and they have no effort to be better



The process of analyzing survey data





Understanding & Collecting Data

Split the data into 2 parts to simplify the analysis process



- **1 = Strongly DisAgree , 4 = Strongly Agree**

N	Question	Scale
1	Sebagian besar hidup saya dilalui untuk melakukan hal yang bermakna	1 (Sangat Tidak Sesuai) - 4 (Sangat Sesuai)
2	Dengan berusaha keras, saya hampir selalu bisa mencapai tujuan	1 (Sangat Tidak Sesuai) - 4 (Sangat Sesuai)
3	Saya tidak ingin mengubah aktivitas rutin yang biasa saya lakukan	1 (Sangat Tidak Sesuai) - 4 (Sangat Sesuai)
4	Saya merasa bahwa terkadang hidup saya tidak bermakna	1 (Sangat Tidak Sesuai) - 4 (Sangat Sesuai)
5	Melakukan hal yang berbeda dari biasanya merupakan hal yang menarik bagi saya	1 (Sangat Tidak Sesuai) - 4 (Sangat Sesuai)

- **1 = Strongly Agree , 4 = Strongly DisAgree**

N	Question	Scale
28	Saya tidak perlu menentukan pekerjaan saya nantinya, karena masa depan itu tidak pasti.	1 (Sangat Setuju) - 4 (Sangat Tidak Setuju)
29	Saya tidak tahu banyak mengenai persyaratan kerja.	1 (Sangat Setuju) - 4 (Sangat Tidak Setuju)
30	Saya memiliki terlalu banyak minat, sehingga saya bingung jika harus memilih salah satu.	1 (Sangat Setuju) - 4 (Sangat Tidak Setuju)
31	Memilih pekerjaan adalah sesuatu yang harus saya lakukan sendiri.	1 (Sangat Setuju) - 4 (Sangat Tidak Setuju)
32	Saya tidak terlalu peduli soal pekerjaan yang akan saya lakukan di masa depan.	1 (Sangat Setuju) - 4 (Sangat Tidak Setuju)



Understanding & Collecting Data

Make data into the same category for Analyzing



Sebagian be	Dengan ber	Saya tidak in	Saya merasa	Melakukan	Sesuatu terj	Saya menan	Tidak terdap	Saya menikr	Hampir sepa
SA	SA	A	D	SA	A	A	D	SA	A
SA	SA	D	SD	SA	SA	SA	SD	SA	SA
A	A	A	D	SA	SA	SA	D	SA	SA
SA	A	A	D	A	A	A	D	A	SA
SA	A	A	D	SA	SA	A	D	A	SA
SA	SD	SA	SD	SA	SA	SA	SD	SA	SA

- 1 = Strongly DisAgree
- 2 = DisAgree
- 3 = Agree
- 4 = Strongly Agree

Sebagian bes	Dengan beru	Saya tidak in	Saya merasa	Melakukan h	Sesuatu terja	Saya menant	Tidak terdap	Saya menikm	Hampir sepa
4	4	2	2	3	3	3	1	4	4
3	4	2	2	4	4	4	1	4	3
3	4	2	2	4	3	2	4	3	2
2	3	2	2	3	3	2	2	2	1
3	4	3	2	3	4	3	2	4	4



Understanding & Collecting Data

Classify each question in which category Using Excel



>> INT = INTROVERT

No	Question	Simbol
38	Saya tidak ingin mengubah aktivitas rutin yang biasa saya lakukan	INT
39	Saya menyukai kegiatan sehari-hari yang tidak terdapat berbagai perubahan	INT11
40	Saya merasa sedikit tertarik dengan program ini	INT15
41	Dalam memilih karir, saya harus memperhatikan pikiran dan perasaan keluarga saya.	INT27
42	Saya tidak tahu banyak mengenai persyaratan kerja.	INT29
43	Memilih pekerjaan adalah sesuatu yang harus saya lakukan sendiri.	INT31
44	Setiap orang berkata hal yang berbeda-beda, sehingga saya menjadi bingung akan pekerjaan yang saya akan pilih.	INT34
45	Saya menantikan kegiatan rutin sehari-hari	INT4
46	Saya merasa terganggu jika hari saya tidak berjalan sesuai rutinitas saya	INT8
47	Terserah pada diri sendiri bagaimana saya memutuskan hidup saya kedepannya	INT9



Understanding & Collecting Data

Classify each question in which category Using Excel



>> EXT = EXTROVERT

No	Question	Simbol
23	Melakukan hal yang berbeda dari biasanya merupakan hal yang menarik bagi saya	EXT
24	Saya menceritakan kepada teman saya bahwa program di sini merupakan program yang bagus untuk diikuti	EXT14
25	Saya merasa senang bahwa saya memilih program ini dibandingkan dengan program yang lain	EXT20
26	Saya peduli dengan keberlangsungan program ini	EXT22
27	Saya ikut program ini hanya untuk coba-coba	EXT24
28	Jika memiliki keraguan mengenai apa yang saya ingin lakukan, saya akan bertanya kepada orangtua atau teman untuk mendapat saran.	EXT25
29	Saat memilih karir, saya akan bertanya kepada orang lain untuk membantu saya.	EXT26
30	Saya memiliki terlalu banyak minat, sehingga saya bingung jika harus memilih salah satu.	EXT30
31	Saya selalu merubah pilihan jalur karir saya.	EXT37
32	Saya tidak tahu kelas apa yang harus saya pilih.	EXT39
33	Saya akan memilih karir saya tanpa memperdulikan perasaan orang lain.	EXT40
34	Saya tidak tahu apakah perencanaan karir saya realistis atau tidak.	EXT41
35	Banyak hal yang harus dipertimbangkan dalam memilih sebuah pekerjaan, maka akan sulit untuk membuat keputusan karir.	EXT42
36	Saya menikmati tantangan saat saya harus melakukan beberapa hal dalam satu waktu	EXT6
37	Hampir sepanjang waktu, hidup terasa sangat menarik bagi saya	EXT7



Understanding & Collecting Data

Classify each question in which category Using Excel



>> CR = CAREER READY

No	Question	Simbol
14	Sebagian besar hidup saya dilalui untuk melakukan hal yang bermakna	CR
15	Keputusan saya membawa perubahan signifikan pada hasil yang akhirnya terjadi	CR12
16	Saya mau berusaha semaksimal mungkin untuk mengikuti program ini	CR13
17	Saya akan mengerjakan berbagai tugas yang diberikan agar dapat sukses di dalam program	CR16
18	Saya merasa bahwa program yang ditawarkan sesuai dengan minat saya	CR17
19	Program ini membuat saya berusaha untuk menunjukkan performa maksimal	CR18
20	Dengan berusaha keras, saya hampir selalu bisa mencapai tujuan	CR2
21	Menurut saya, program ini merupakan tempat terbaik untuk meningkatkan skill saya	CR23
22	Sesuatu terjadi dalam hidup tergantung pada apa yang saya lakukan	CR3



Understanding & Collecting Data

Classify each question in which category Using Excel



>> **CNR = CAREER NOT READY**

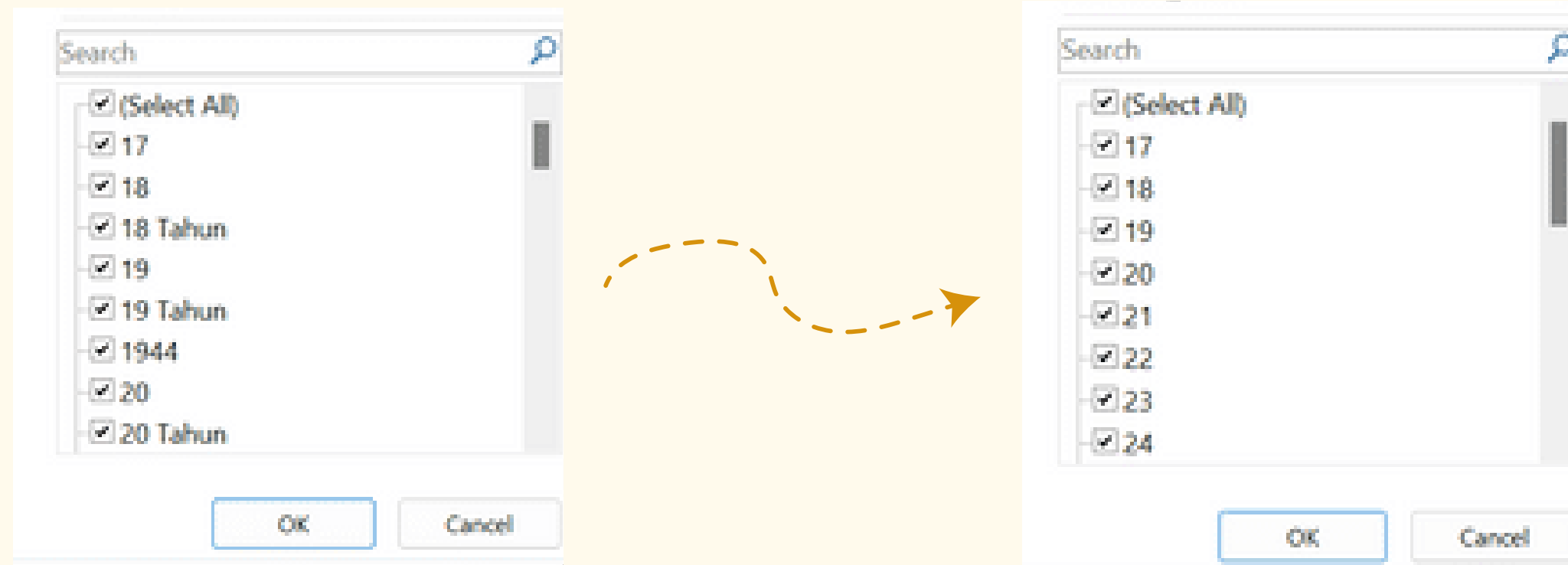
No	Question	Simbol
1	Saya merasa bahwa terkadang hidup saya tidak bermakna	CNR
2	Secara umum, hidup membosankan bagi saya	CNR10
3	Kesulitan yang saya hadapi membuat saya memilih untuk berhenti	CNR19
4	Tidak terlalu banyak hal menarik yang saya peroleh dari program ini	CNR21
5	Saya tidak perlu menentukan pekerjaan saya nantinya, karena masa depan itu tidak pasti.	CNR28
6	Saya tidak terlalu peduli soal pekerjaan yang akan saya lakukan di masa depan.	CNR32
7	Saya tidak tahu bagaimana saya akan mendapatkan pekerjaan yang saya impikan.	CNR33
8	Saya jarang memikirkan pekerjaan yang ingin dilakukan.	CNR35
9	Saya memiliki kesulitan mempersiapkan diri untuk pekerjaan yang saya ingin lakukan.	CNR36
10	Saya tidak akan khawatir dalam memilih pekerjaan selama saya belum lulus kuliah.	CNR38
11	Saya benar-benar tidak bisa menemukan sebuah pekerjaan yang menarik bagi saya.	CNR43
12	Saya tidak paham bagaimana orang lain dapat begitu yakin dengan apa yang ingin mereka lakukan.	CNR44
13	Tidak terdapat banyak hal yang bisa saya lakukan untuk mempengaruhi masa depan saya	CNR5



Data Cleansing



- Change the writing model that doesn't fit
- Handle Missing Value and Change Type Data



```
[ ] ## handle missing value dengan fill median (numeric)  
data = data.fillna(data.iloc[:, [32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]].median())
```



```
## Mengubah float ke int
```

```
data.iloc[:, [32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]]  
= data.iloc[:, [32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]].astype(int)
```

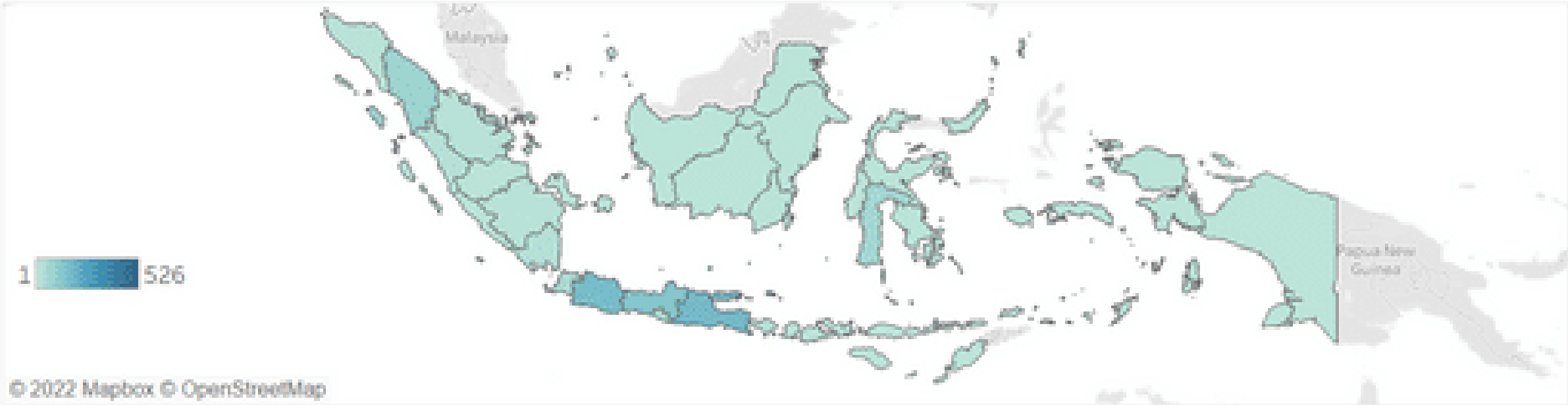



Exploratory Data Analysis

You can access the dashboard at the link :

<https://public.tableau.com/app/profile/claryta.putri.dedyana/viz/FinalProjectMESOngoing/DashboardEDA?publish=yes>

Map User Distribution



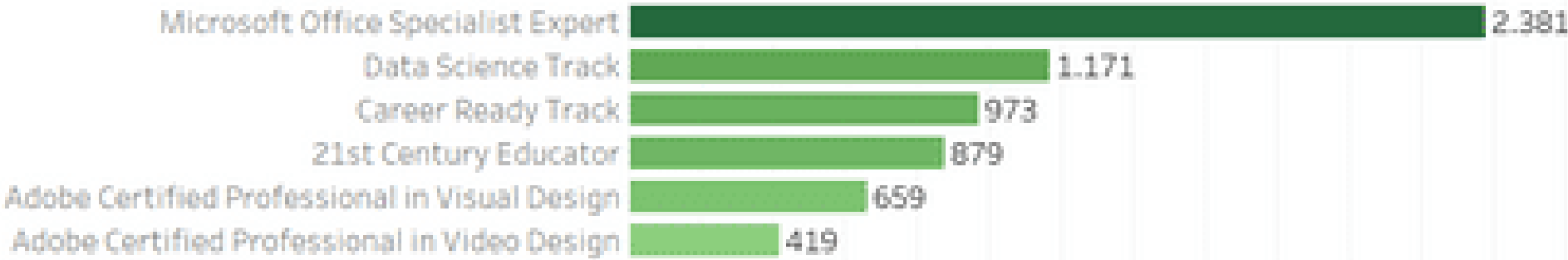
EXPLORATORY
DATA ANALYSIS

- ☒ (All)
- ☒ 21st Century Educator
- ☒ Adobe Certified Professional in Vi...
- ☒ Adobe Certified Professional in Vi...
- ☒ Career Ready Track
- ☒ Career Ready Track, Ms.Office, da...
- ☒ Data Science Track
- ☒ Microsoft Office Specialist Expert

User Perfomance by Age

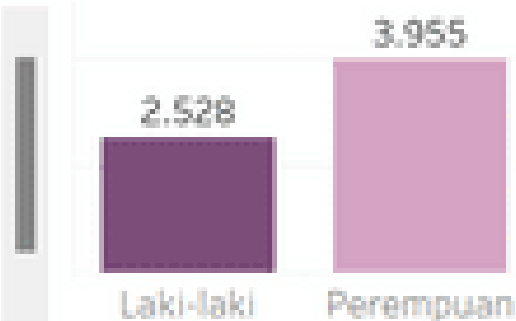
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	33	34	35	36	38	39
Microsoft Office Specialist Expert	4	105	880	982	323	50	20	8		1	3		3		1					
Data Science Track	1	55	451	461	146	30	14		4	1	2	1	1	1			1	1		
Career Ready Track	2	37	382	402	113	22	6	4	2	1	1	1								
21st Century Educator	2	44	316	351	130	17	6	6	2	1	2				1				1	
Adobe Certified Professional in Visual Design	4	37	244	243	91	31	3	3		1	2									
Adobe Certified Professional in Video Design		15	165	165	42	16	5	4	2	2			1			1				

Program Category Interest

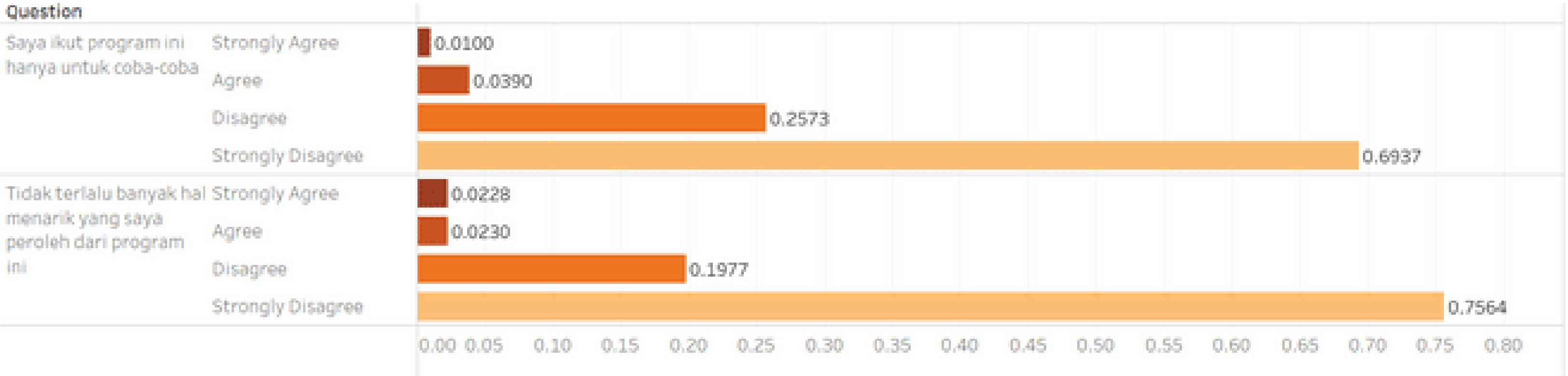
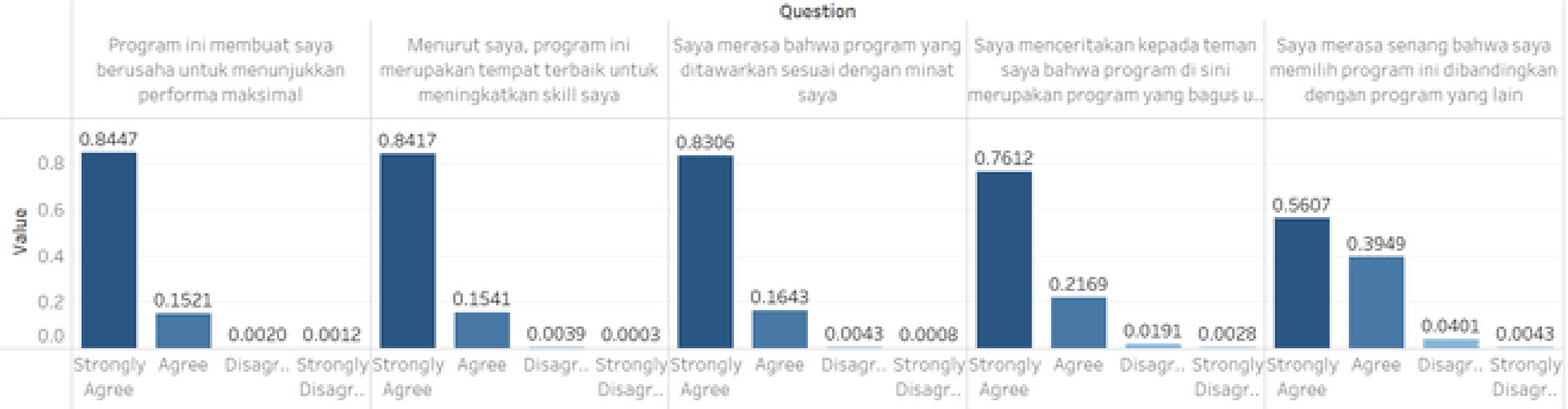


6483
Total Users
Response

User



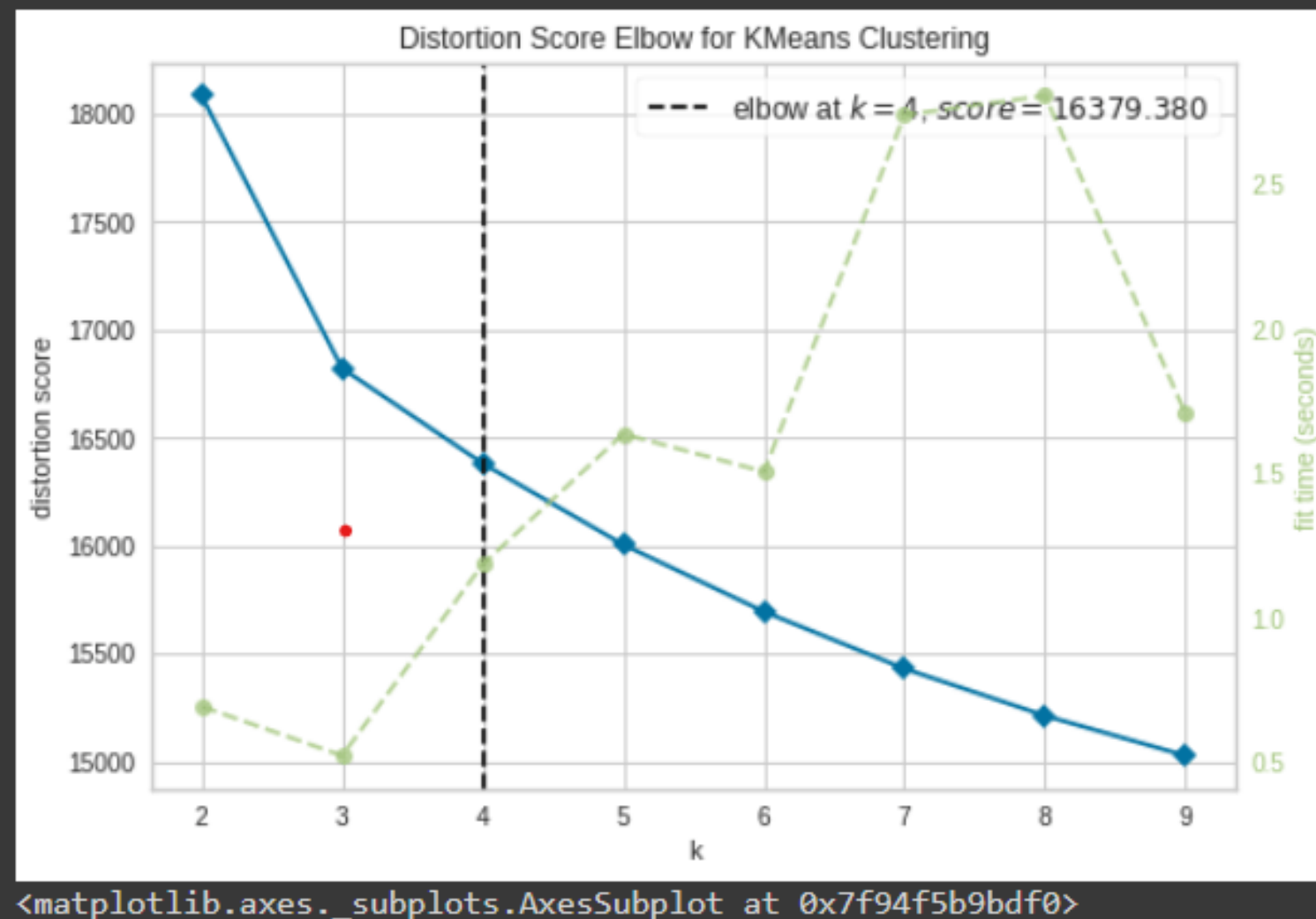
PROGRAM PERFORMANCE MES



Modelling python™



```
# Visualize the elbow
from sklearn.cluster import KMeans
from yellowbrick.cluster import KElbowVisualizer
kmeans = KMeans()
visualizer = KElbowVisualizer(kmeans, k=(2,10))
visualizer.fit(df)
visualizer.poof()
```



As you can see 4 clusters look optimum for the data set and we already know this project is to identify 4 different personalities (Extrovert, Introvert, Career Ready, Career Not Ready)

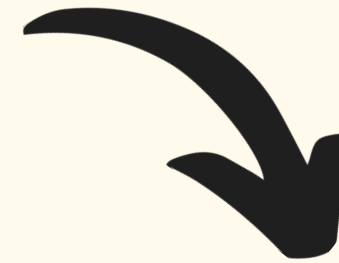
Modelling python™

Get prediction analysis cluster for each user

```
# I define 4 clusters and fit my model  
kmeans = KMeans(n_clusters=4)  
k_fit = kmeans.fit(df_model)
```



```
# Predicting the Clusters  
predictions = k_fit.labels_  
df_model['Clusters'] = predictions  
df_model
```



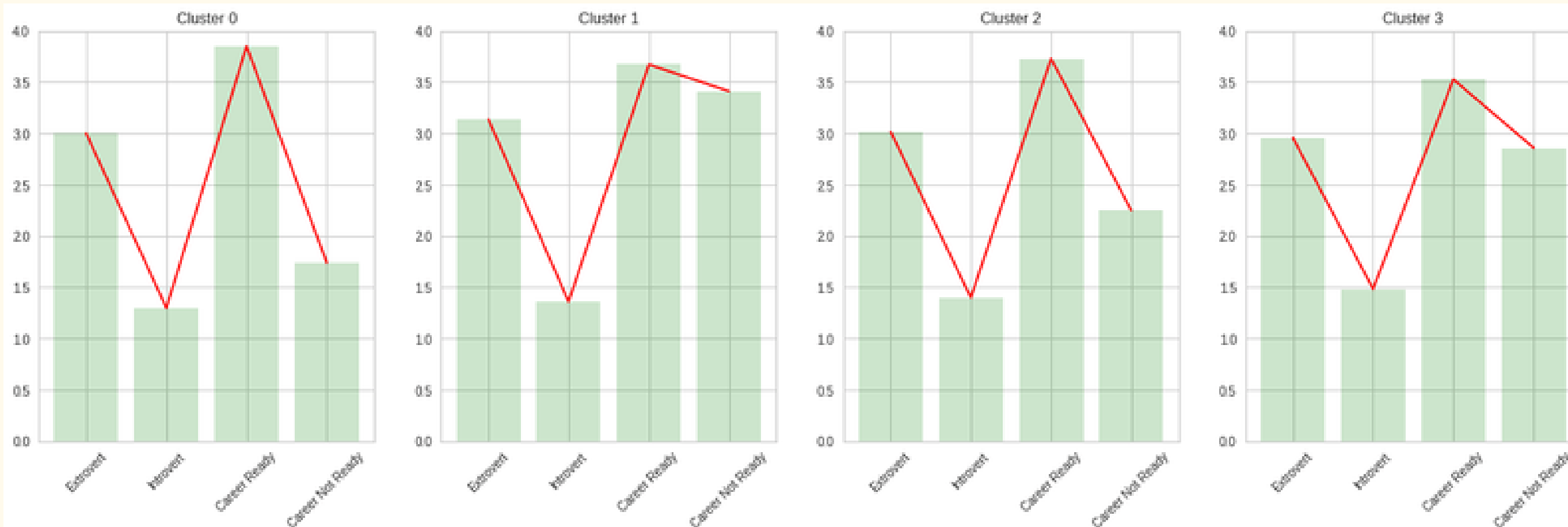
df_model																					
	CR	CR2	CNR	CNR3	EXT	CR4	INT	CNR5	EXT6	EXT7	...	EXT37	CNR38	CNR39	EXT40	CNR41	INT42	CNR43	CNR44	Clusters	ID
0	4	4	2	2	3	3	3	1	4	4	...	3	1	1	2	3	3	1	4	2	1001
1	3	4	2	2	4	4	4	1	4	3	...	1	1	1	1	2	3	1	1	0	1002
2	3	4	2	2	4	3	2	4	3	2	...	1	1	2	2	4	4	1	3	3	1003
3	2	3	2	2	3	3	2	2	2	1	...	2	2	3	2	4	3	3	4	3	1004
4	3	4	3	2	3	4	3	2	4	4	...	4	4	1	4	1	3	1	2	2	1005
...
6477	3	4	2	2	4	3	2	3	4	3	...	3	2	2	2	2	2	2	3	3	7480
6478	3	3	4	2	4	3	3	3	2	4	...	3	4	3	3	3	2	4	1	1	7481

Modelling python™

clusters	0	1	2	3
Extrovert	3.000642	3.137420	3.017116	2.957110
Introvert	1.298340	1.360784	1.399497	1.483108
Career Ready	3.849263	3.671484	3.730476	3.526518
Career Not Ready	1.739523	3.411298	2.249329	2.859896

```
df_model.Clusters.value_counts()
```

```
2    2292
0    1797
3    1433
1     960
Name: Clusters, dtype: int64
```



Most types of user behavior that take the program in MyEduSolve : Extrovert and Career Ready

Modelling python™

Visualizing The Cluster Prediction

Visualizing the Cluster Predictions

```
[ ] # In order to visualize in 2D graph I will use PCA
    from sklearn.decomposition import PCA
```

```
pca = PCA(n_components=2)
pca_fit = pca.fit_transform(df_model)
```

```
df_pca = pd.DataFrame(data=pca_fit, columns=['PCA1', 'PCA2'])
df_pca['Clusters'] = predictions
df_pca.head()
```

	PCA1	PCA2	Clusters
0	-0.372305	-0.996260	2
1	-3.015519	-0.326268	0
2	0.668732	-3.737934	3
3	5.290689	-1.956160	3
4	0.666479	-0.681728	2

```
[ ] plt.figure(figsize=(10,10))
    sns.scatterplot(data=df_pca, x='PCA1', y='PCA2', hue='Clusters', palette='Set2', alpha=0.8)
    plt.title('Personality Clusters after PCA');
```



●●● Conclusion



Those who register for the MyEduSolve program are more dominated by career ready and extrovert personalities

The MES team can prepare technical matters such as learning media, the number of mentors needed, and the number of PICs who accompany their students

increase the reception quota on programs that have high demand (Microsoft Office Specialist Expert)

Learning methods can be adapted to the characteristics of the participants' paths with the aim of maximizing the learning process and determining an effective way of learning



**THANK
YOU**

