



CAPSTONE PROJECT

**"Personality Clustering"
Based on Big 5 Personality**

By Group 11

6

INTRODUCTION



Alifia



Diyman



Arival



Jasmine



Tasya



META DATA

1

Data Understanding & Exploration

2

Data Cleansing

3

Modeling

4

Visualization

Clustering Personality based on Big 5 Personality

1. Openness to Experience (Terbuka terhadap Hal-hal baru)

“Saya tidak ingin mengubah aktifitas rutin yang biasa saya lakukan”

“Melakukan hal yang berbeda dari biasanya merupakan hal yang menarik bagi saya”

“Saya menyukai kegiatan sehari-hari yang tidak terdapat berbagai perubahan”



Clustering Personality based on Big 5 Personality



2. Conscientiousness (Sifat Berhati-hati)

"Saya akan mengerjakan berbagai tugas yang diberikan agar dapat sukses di dalam program"

"Saat memilih karir, saya akan bertanya kepada orang lain untuk membantu saya"

"Banyak hal yang harus dipertimbangkan dalam memilih sebuah pekerjaan, maka akan sulit untuk membuat keputusan karir"



Clustering Personality based on Big 5 Personality

3. Extraversion (Ekstraversi)

“Saya menantikan kegiatan rutin sehari hari”

“Saya menceritakan kepada teman saya bahwa program disini merupakan program yang bagus untuk diikuti”

“secara umum, hidup membosankan bagi saya”



Clustering Personality based on Big 5 Personality

4. Agreeableness (Mudah Akur atau Mudah Bersepakat)

“Jika memiliki keraguan mengenai apa yang ingin saya lakukan, Saya akan bertanya kepada orangtua atau teman untuk mendapat saran”

“Dalam memilih karir, saya harus memperhatikan pikiran dan perasaan keluarga saya”

“Saya akan memilih karir saya tanpa memperdulikan perasaan orang lain”



DATA UNDERSTANDING

Clustering Personality based on Big 5 Personality

5. Neuroticism (Neurotisme)

"Dengan berusaha keras, saya hampir selalu bisa mencapai tujuan"

"Saya menikmati tantangan saat saya harus melakukan beberapa hal dalam satu waktu"

"Kesulitan yang saya hadapi membuat saya memilih untuk berhenti"



PEMBOBOTAN

Pembobotan		
Nilai	Ekspetasi Score Tinggi	Ekspetasi Score Rendah
1	1	4
2	2	3
3	3	2
4	4	1

Pembobotan kesesuaian skala
nilai jawaban dengan point
pertanyaan yang diinginkan

Keterangan Pertanyaan	
	Ekspetasi Score Rendah
	Ekspetasi Score Tinggi

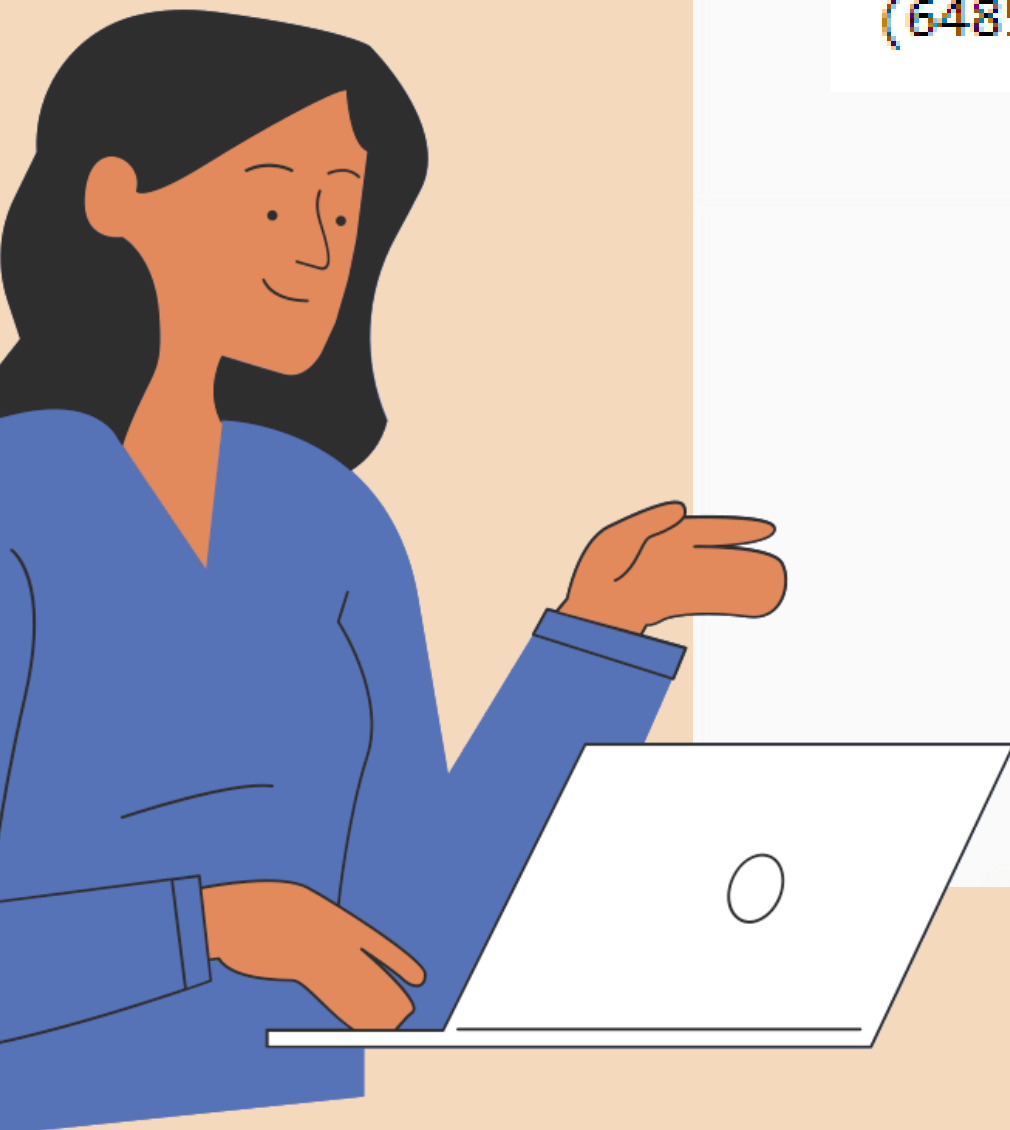
DATA EXPLORATION



```
data.shape  
(6485, 52)
```

```
In [24]: data.columns
```

```
Out[24]: Index(['ID', 'Usia', 'Jenis Kelamin', 'Tempat tinggal saat ini',  
               'Pilihan pertama program yang diminati', 'Pertanyaan 1 ',  
               'Pertanyaan 2 ', 'pertanyaan 3', 'Pertanyaan 4', 'Pertanyaan 5',  
               'Pertanyaan 6', 'Pertanyaan 7', 'Pertanyaan 8', 'Pertanyaan 9',  
               'Pertanyaan 10', 'Pertanyaan 11', 'Pertanyaan 12', 'Pertanyaan 13',  
               'Pertanyaan 14', 'Pertanyaan 15', 'Pertanyaan 16', 'Pertanyaan 17',  
               'Pertanyaan 18', 'Pertanyaan 19', 'Pertanyaan 20', 'Pertanyaan 21',  
               'Pertanyaan 22', 'Pertanyaan 23', 'Pertanyaan 24', 'Pertanyaan 25',  
               'Pertanyaan 26', 'Pertanyaan 27', 'Pertanyaan 28', 'Pertanyaan 29',  
               'Pertanyaan 30', 'Pertanyaan 31', 'Pertanyaan 32', 'Pertanyaan 33',  
               'Pertanyaan 34', 'Pertanyaan 35', 'Pertanyaan 36', 'Pertanyaan 37',  
               'Pertanyaan 38', 'Pertanyaan 39', 'Pertanyaan 40', 'Pertanyaan 41',  
               'Pertanyaan 42', 'Pertanyaan 43', 'Pertanyaan 44', 'Pertanyaan 45',  
               'Pertanyaan 46', 'Pertanyaan 47'],  
              dtype='object')
```



DATA EXPLORATION



```
data['Pilihan pertama program yang diminati'].unique()

array(['Data Science Track', 'Career Ready Track',
      'Adobe Certified Professional in Visual Design',
      'Microsoft Office Specialist Expert', '21st Century Educator',
      'Adobe Certified Professional in Video Design',
      'Career Ready Track, Ms.Office, data science'], dtype=object)
```

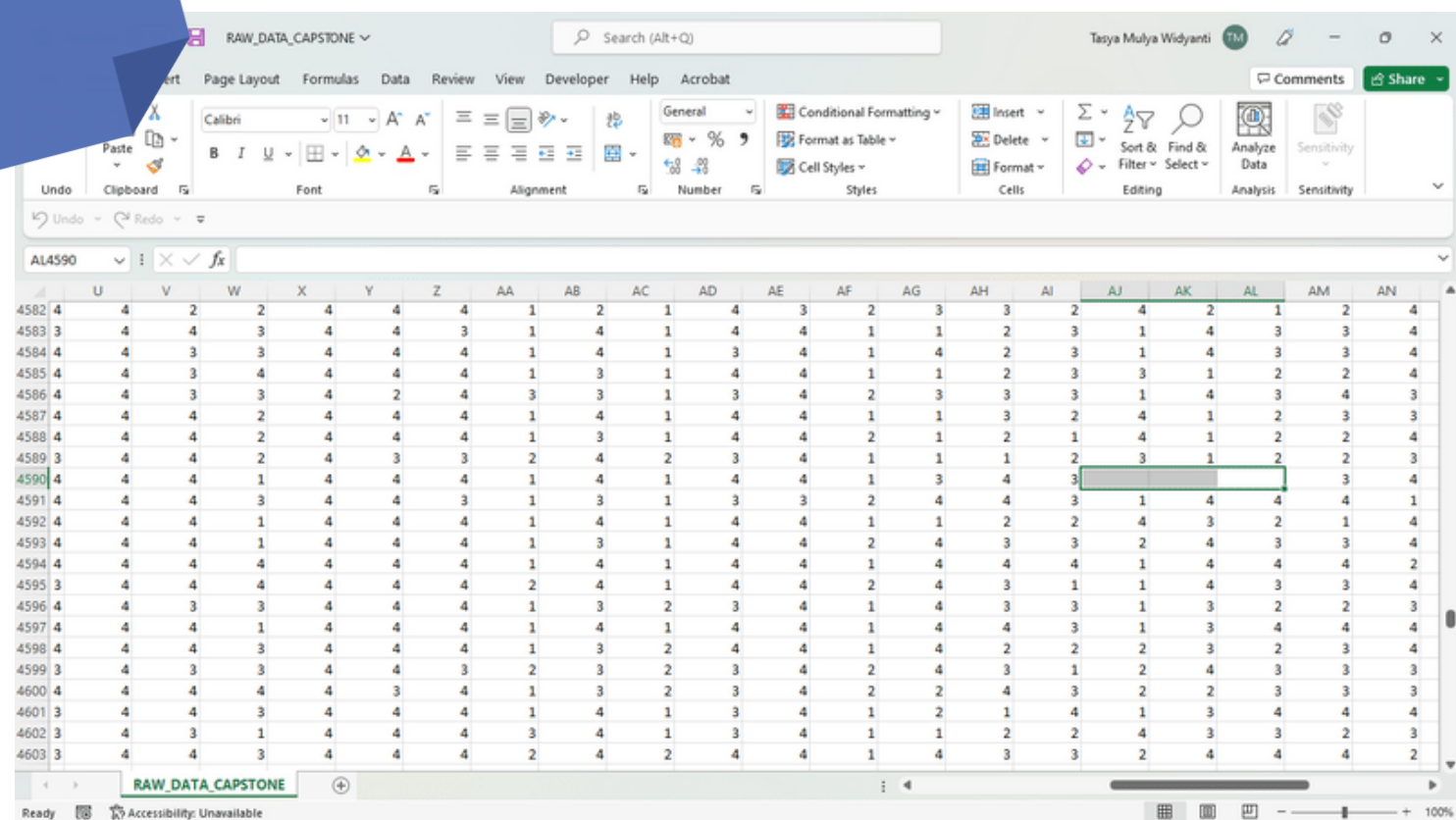
```
data['Pilihan pertama program yang diminati'].value_counts()
```

Microsoft Office Specialist Expert	2381
Data Science Track	1171
Career Ready Track	973
21st Century Educator	880
Adobe Certified Professional in Visual Design	659
Adobe Certified Professional in Video Design	419
Career Ready Track, Ms.Office, data science	1
Name: Pilihan pertama program yang diminati, dtype: int64	



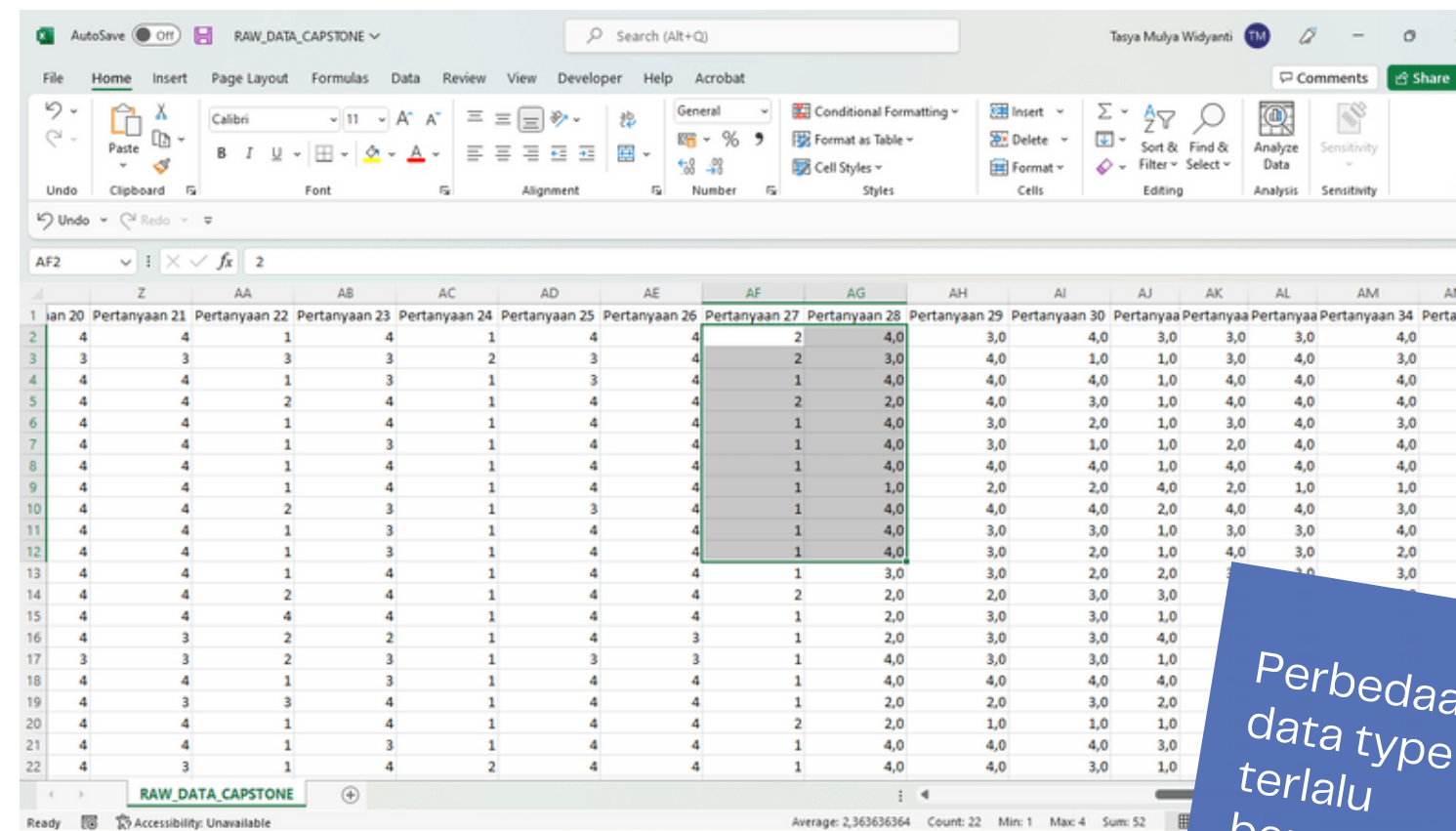
DATA CLEANSING

Missing value
diatasi dengan
"Modus"



	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
4582	4	4	2	2	4	4	4	1	2	1	4	3	2	3	2	4	2	1	2	4
4583	3	4	4	3	4	4	3	1	4	1	4	4	1	1	2	3	1	4	3	4
4584	4	4	3	3	4	4	4	1	4	1	3	4	1	4	2	3	1	4	3	4
4585	4	4	3	4	4	4	4	1	3	1	4	4	1	1	2	3	3	1	2	4
4586	4	4	3	3	4	2	4	3	3	1	3	4	2	3	3	3	1	4	3	4
4587	4	4	4	2	4	4	4	1	4	1	4	4	1	1	3	2	4	1	2	3
4588	4	4	4	2	4	4	4	1	3	1	4	4	2	1	2	1	4	1	2	4
4589	4	4	4	2	4	3	3	2	4	2	3	4	1	1	1	2	3	1	2	3
4590	4	4	4	1	4	4	4	1	4	1	4	4	1	3	4	3				
4591	4	4	4	3	4	4	3	1	3	1	3	3	2	4	4	3	1	4	4	1
4592	4	4	4	1	4	4	4	1	4	1	4	4	1	1	2	2	4	3	2	1
4593	4	4	4	1	4	4	4	1	3	1	4	4	2	4	3	3	2	4	3	4
4594	4	4	4	4	4	4	4	1	4	1	4	4	1	4	4	4	1	4	4	2
4595	3	4	4	4	4	4	4	2	4	1	4	4	2	4	3	1	1	4	3	4
4596	4	4	3	3	4	4	4	1	3	2	3	4	1	4	3	3	1	3	2	3
4597	4	4	4	1	4	4	4	1	4	1	4	4	1	4	4	3	1	3	4	4
4598	4	4	4	3	4	4	4	1	3	2	4	4	1	4	2	2	2	3	2	4
4599	3	4	3	3	4	4	3	2	3	2	3	4	2	4	3	1	2	4	3	3
4600	4	4	4	4	4	3	4	1	3	2	3	4	2	2	4	3	2	2	3	3
4601	3	4	4	3	4	4	4	1	4	1	3	4	1	2	1	4	1	3	4	4
4602	3	4	3	1	4	4	4	3	4	1	3	4	1	1	2	2	4	3	3	2
4603	3	4	4	3	4	4	4	2	4	2	4	4	1	4	3	3	2	4	4	2

missing value & data type checking



	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
1	Pertanyaan 20	Pertanyaan 21	Pertanyaan 22	Pertanyaan 23	Pertanyaan 24	Pertanyaan 25	Pertanyaan 26	Pertanyaan 27	Pertanyaan 28	Pertanyaan 29	Pertanyaan 30	Pertanyaan 31	Pertanyaan 32	Pertanyaan 33	Pertanyaan 34
2	4	4	1	4	1	4	4	2	4,0	3,0	4,0	3,0	3,0	4,0	4,0
3	3	3	3	3	2	3	4	2	3,0	4,0	1,0	3,0	4,0	3,0	3,0
4	4	4	1	3	1	3	4	1	4,0	4,0	4,0	1,0	4,0	4,0	4,0
5	4	4	2	4	1	4	4	2	2,0	4,0	3,0	1,0	4,0	4,0	4,0
6	4	4	1	4	1	4	4	1	4,0	3,0	2,0	1,0	3,0	4,0	3,0
7	4	4	1	3	1	4	4	1	4,0	3,0	1,0	1,0	2,0	4,0	4,0
8	4	4	1	4	1	4	4	1	4,0	4,0	4,0	1,0	4,0	4,0	4,0
9	4	4	1	4	1	4	4	1	1,0	2,0	2,0	4,0	2,0	1,0	1,0
10	4	4	2	3	1	3	4	1	4,0	4,0	4,0	2,0	4,0	4,0	3,0
11	4	4	1	3	1	4	4	1	4,0	3,0	3,0	1,0	3,0	3,0	4,0
12	4	4	1	3	1	4	4	1	4,0	3,0	2,0	1,0	4,0	3,0	2,0
13	4	4	1	4	1	4	4	1	3,0	3,0	2,0	2,0	3,0	3,0	3,0
14	4	4	2	4	1	4	4	2	2,0	2,0	3,0	3,0			
15	4	4	4	4	1	4	4	1	2,0	3,0	3,0	1,0			
16	4	3	2	2	1	4	3	1	2,0	3,0	3,0	4,0			
17	3	3	2	3	1	3	3	1	4,0	3,0	3,0	1,0			
18	4	4	1	3	1	4	4	1	4,0	4,0	4,0	4,0			
19	4	3	3	4	1	4	4	1	2,0	2,0	3,0	2,0			
20	4	4	1	4	1	4	4	2	2,0	1,0	1,0	1,0			
21	4	4	1	3	1	4	4	1	4,0	4,0	4,0	3,0			
22	4	3	1	4	2	4	4	1	4,0	4,0	3,0	1,0			

Perbedaan
data type tidak
terlalu
berpengaruh!

DATA CLEANSING

missing value handling

```
#mengganti missing value dengan menggunakan mode
data['Pertanyaan 30'] = data['Pertanyaan 30'].replace(np.NaN,3)
data['Pertanyaan 31'] = data['Pertanyaan 31'].replace(np.NaN,1)
data['Pertanyaan 32'] = data['Pertanyaan 32'].replace(np.NaN,4)
data['Pertanyaan 33'] = data['Pertanyaan 33'].replace(np.NaN,3)
data['Pertanyaan 34'] = data['Pertanyaan 34'].replace(np.NaN,4)
data['Pertanyaan 35'] = data['Pertanyaan 35'].replace(np.NaN,4)
data['Pertanyaan 36'] = data['Pertanyaan 36'].replace(np.NaN,4)
data['Pertanyaan 37'] = data['Pertanyaan 37'].replace(np.NaN,3)
data['Pertanyaan 38'] = data['Pertanyaan 38'].replace(np.NaN,3)
data['Pertanyaan 39'] = data['Pertanyaan 39'].replace(np.NaN,3)
data['Pertanyaan 40'] = data['Pertanyaan 40'].replace(np.NaN,4)
data['Pertanyaan 41'] = data['Pertanyaan 41'].replace(np.NaN,4)
data['Pertanyaan 42'] = data['Pertanyaan 42'].replace(np.NaN,3)
data['Pertanyaan 43'] = data['Pertanyaan 43'].replace(np.NaN,3)
data['Pertanyaan 44'] = data['Pertanyaan 44'].replace(np.NaN,2)
data['Pertanyaan 45'] = data['Pertanyaan 45'].replace(np.NaN,4)
data['Pertanyaan 46'] = data['Pertanyaan 46'].replace(np.NaN,3)
data['Pertanyaan 47'] = data['Pertanyaan 47'].replace(np.NaN,3)
```

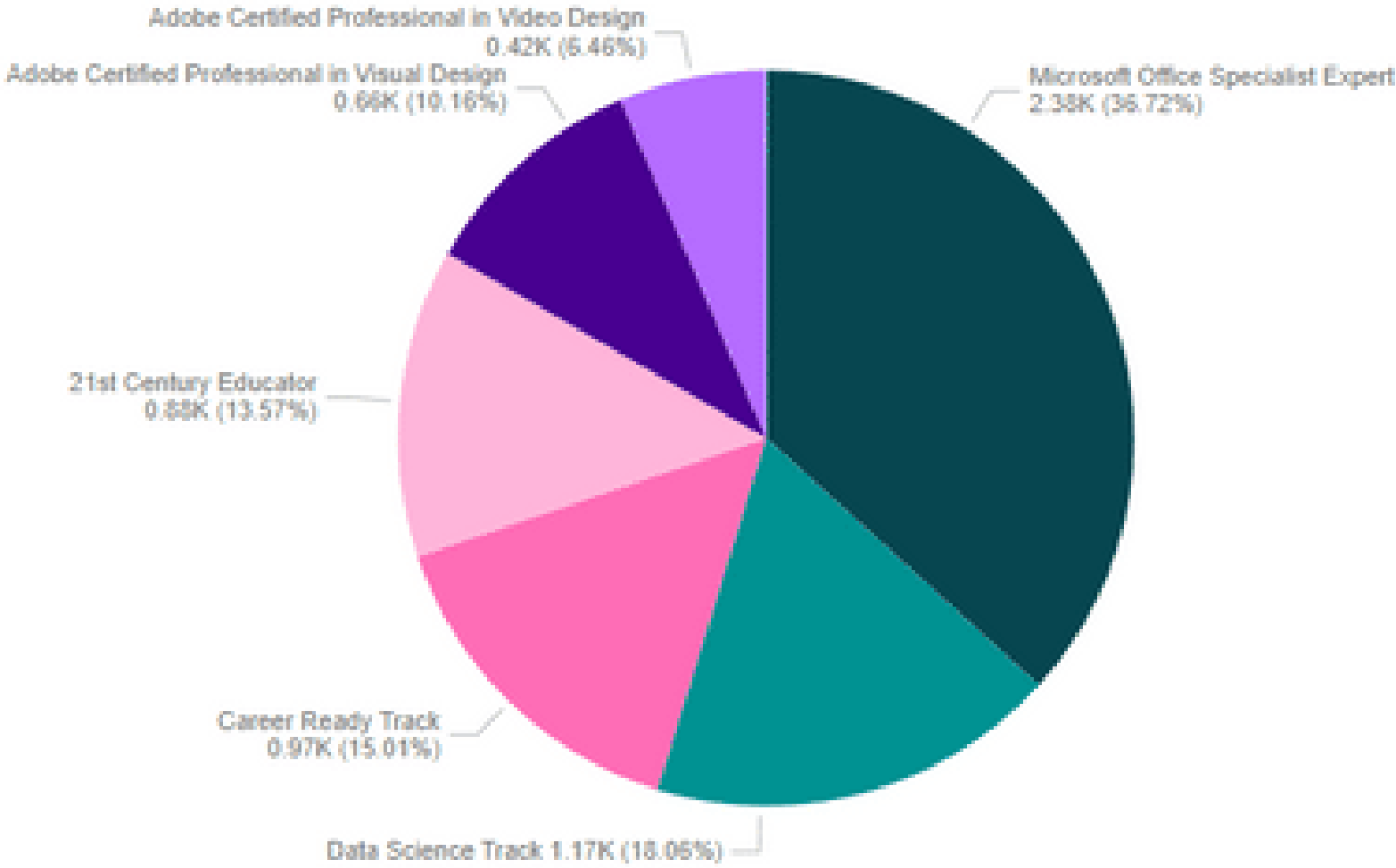
Pengecekan kembali hasil cleaning data

```
#Menghitung missing value
data.isna().sum()
```

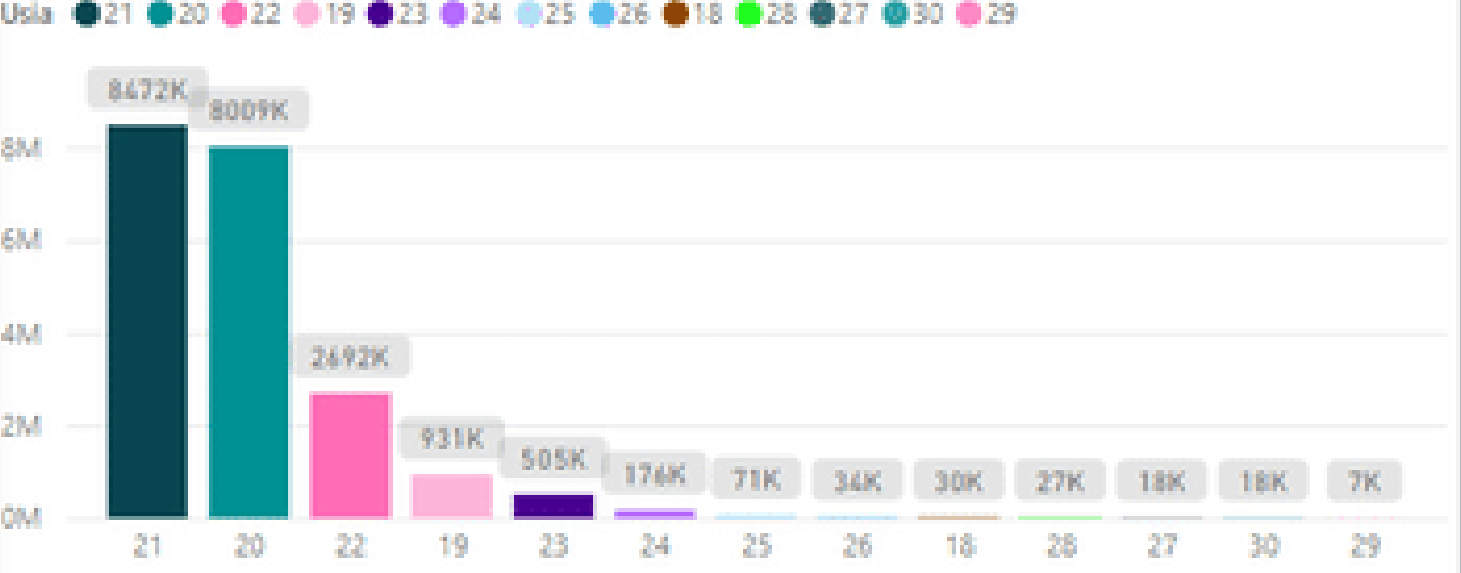
ID	0	
Usia	0	
Jenis Kelamin	0	
Tempat tinggal saat ini	0	
Pilihan pertama program yang diminati	0	Pertanyaan 23
Pertanyaan 1	0	Pertanyaan 24
Pertanyaan 2	0	Pertanyaan 25
pertanyaan 3	0	Pertanyaan 26
Pertanyaan 4	0	Pertanyaan 27
Pertanyaan 5	0	Pertanyaan 28
Pertanyaan 6	0	Pertanyaan 29
Pertanyaan 7	0	Pertanyaan 30
Pertanyaan 8	0	Pertanyaan 31
Pertanyaan 9	0	Pertanyaan 32
Pertanyaan 10	0	Pertanyaan 33
Pertanyaan 11	0	Pertanyaan 34
Pertanyaan 12	0	Pertanyaan 35
Pertanyaan 13	0	Pertanyaan 36
Pertanyaan 14	0	Pertanyaan 37
Pertanyaan 15	0	Pertanyaan 38
Pertanyaan 16	0	Pertanyaan 39
Pertanyaan 17	0	Pertanyaan 40
Pertanyaan 18	0	Pertanyaan 41
Pertanyaan 19	0	Pertanyaan 42
Pertanyaan 20	0	Pertanyaan 43
Pertanyaan 21	0	Pertanyaan 44
Pertanyaan 22	0	Pertanyaan 45
		Pertanyaan 46
		Pertanyaan 47
		dtype: int64

VISUALIZATION

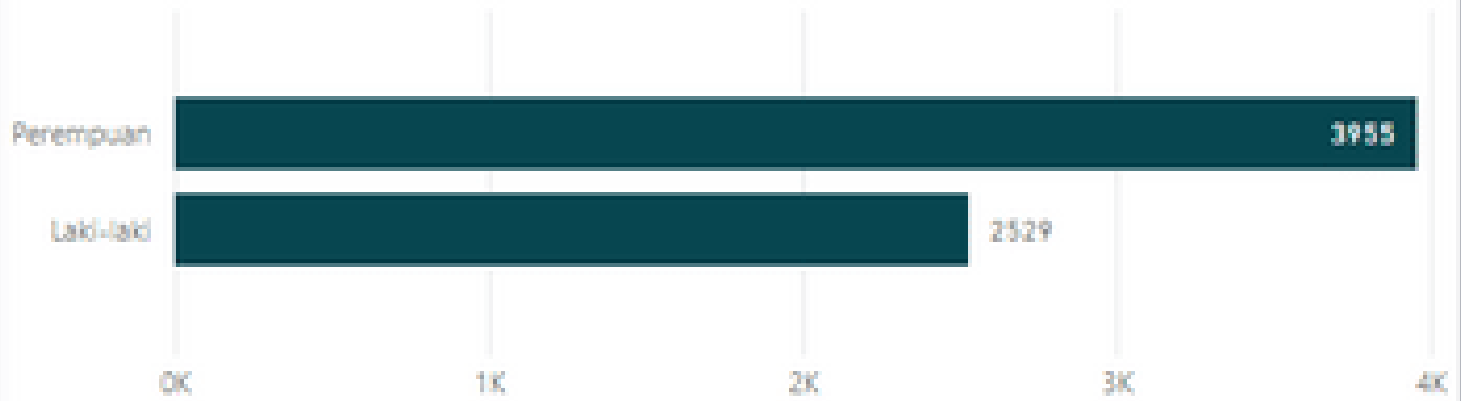
Pilihan pertama program yang diminati



sebaran data usia yang mengikuti program



count of jenis kelamin by ID

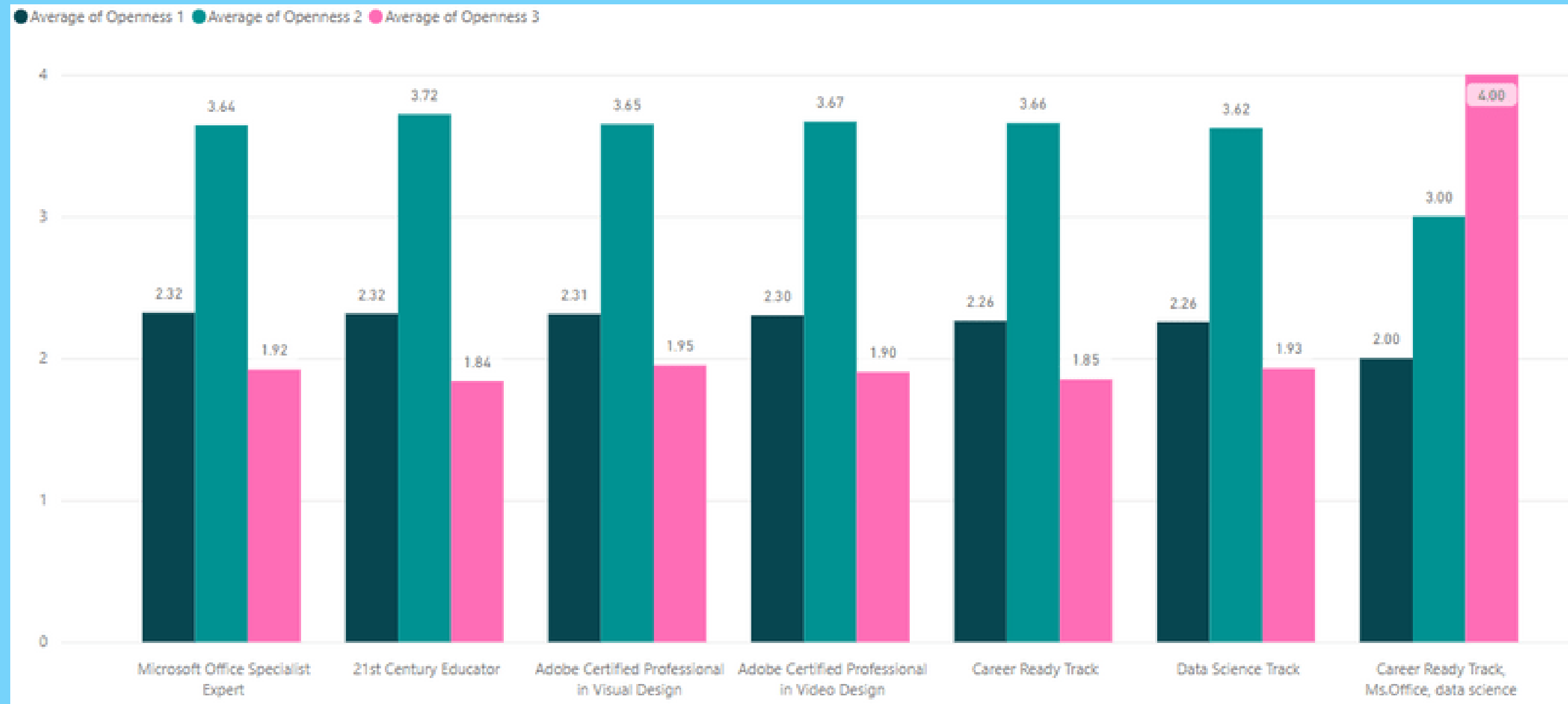


Distribution of customer by City



VISUALIZATION

OPENNESS



Dari visualisasi berikut dapat dilihat bahwa kepribadian dari ke 7 program memiliki kepribadian openness yang tinggi, yang dimana kepribadian ini memiliki sifat yaitu ketertarikannya terhadap hal-hal baru dan keinginan untuk mengetahui serta mempelajari sesuatu yang baru. Karakteristik positif pada Individu yang memiliki dimensi ini cenderung lebih kreatif, Imajinatif, Intelktual, penasaran dan berpikiran luas.

VISUALIZATION

CONSCIENTIOUSNESS



Dari visualisasi berikut dapat dilihat bahwa kepribadian dari ke 7 program memiliki kepribadian conscientiousness yang tinggi, yang dimana kepribadian ini cenderung lebih berhati-hati dalam melakukan suatu tindakan ataupun penuh pertimbangan dalam mengambil sebuah keputusan, mereka juga memiliki disiplin diri yang tinggi dan dapat dipercaya. Karakteristik Positif pada dimensi adalah dapat diandalkan, bertanggung jawab, tekun dan berorientasi pada pencapaian.

VISUALIZATION

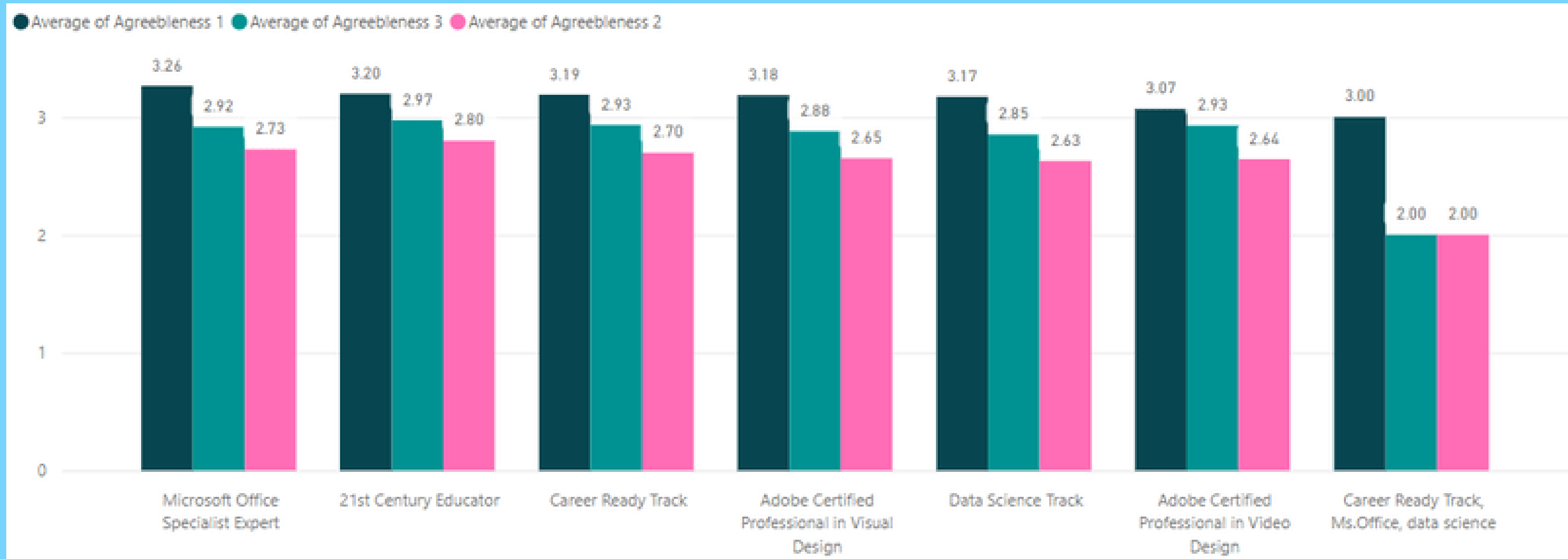
EXTRAVERSION



Dari visualisasi berikut dapat dilihat bahwa kepribadian dari ke 7 program memiliki kepribadian extraversion yang tinggi, yang dimana kepribadian ini senang bergaul, mudah bersosialisasi, hidup berkelompok dan tegas.

VISUALIZATION

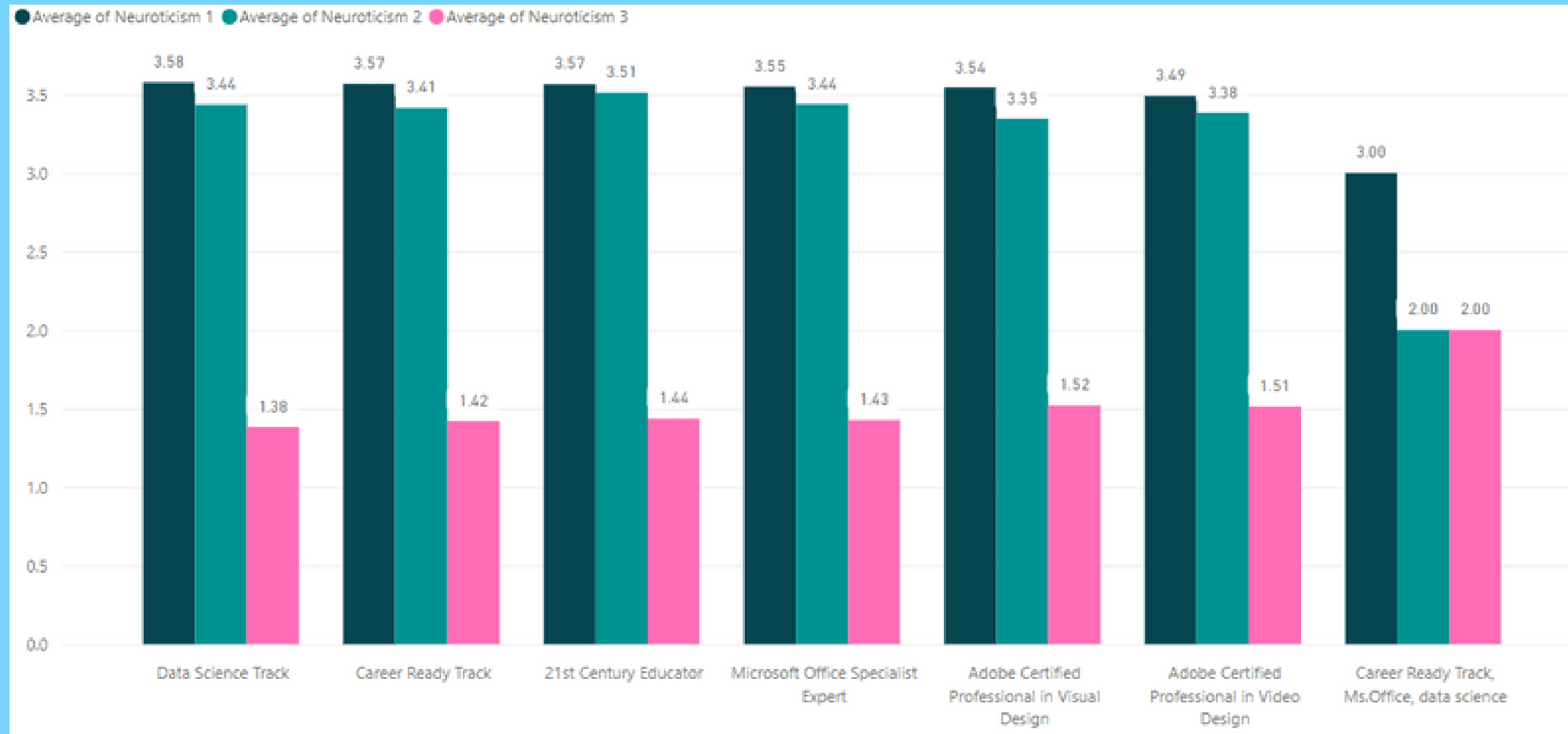
AGREEABLENESS



Dari visualisasi berikut dapat dilihat bahwa kepribadian dari ke 7 program memiliki kepribadian agreeableness yang tinggi, yang dimana kepribadian ini cenderung lebih patuh dengan individu lainnya dan memiliki kepribadian yang ingin menghindari konflik. Karakteristik Positif-nya adalah kooperatif (dapat bekerjasama), penuh kepercayaan, bersifat baik, hangat dan berhati lembut serta suka membantu.

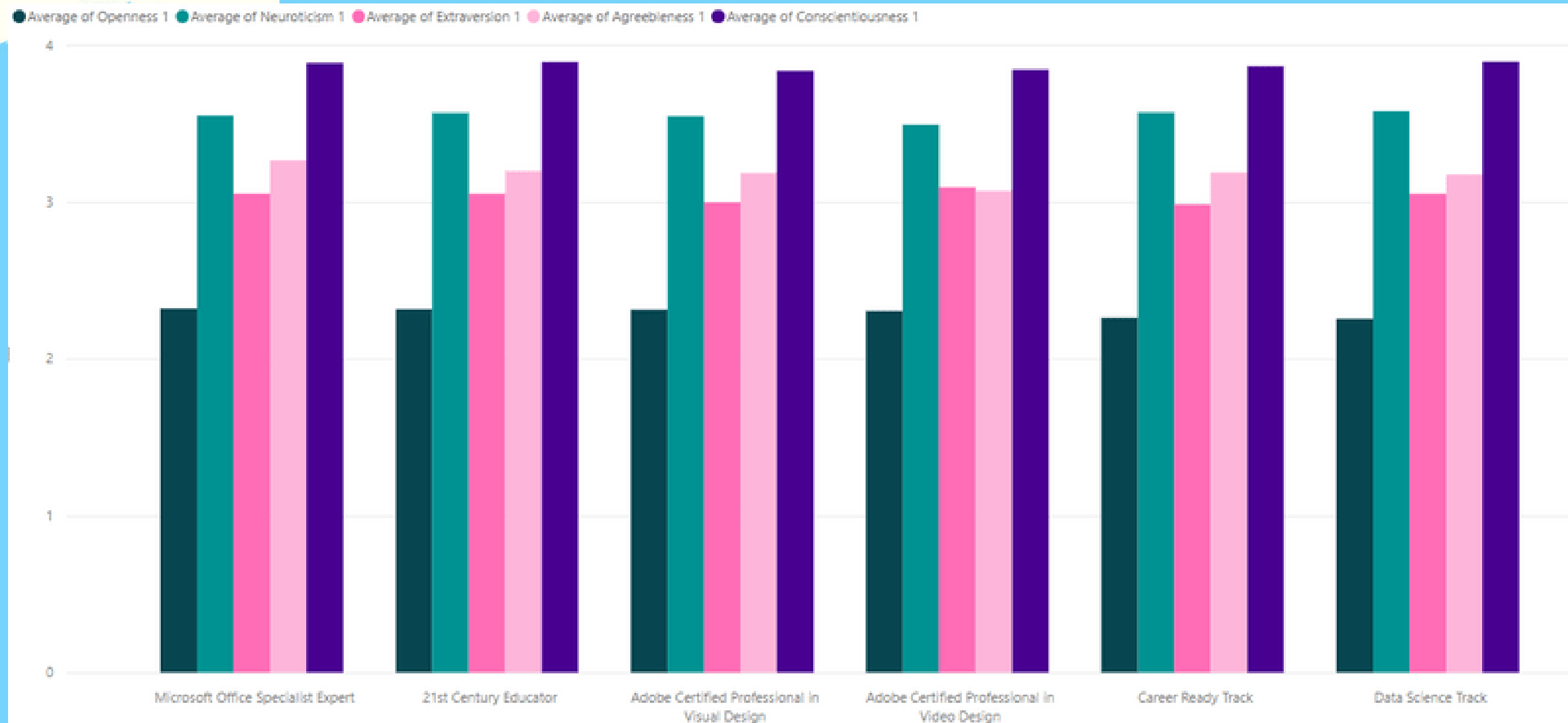
VISUALIZATION

NEUROTICISM



Dari visualisasi berikut dapat dilihat bahwa kepribadian dari ke 7 program memiliki kepribadian neuroticism yang tinggi, yang dimana kepribadian ini merupakan Individu dengan Emosional yang stabil cenderung Tenang saat menghadapi masalah, percaya diri, memiliki pendirian yang teguh.

VISUALIZATION



Berikut hasil visualisasi keseluruhan program dengan kelima personality BIG 5, dapat dilihat bahwa seluruh program memiliki kepribadian yang sama yaitu conscientiousness yang dimana kepribadian ini memiliki sifat yang berhati hati dalam melakukan suatu tindakan dan penuh dengan pertimbangan, disiplin, dapat dipercaya, serta berorientasi pada pencapaian.

Dan yang paling rendah yaitu kepribadian Openess yang berarti peserta kurang memiliki sifat yang kreatif, Imajinatif, Intelektual, penasaran dan berpikiran luas.



Conclusion

Dari hasil analisis kami dapat kami simpulkan bahwa seluruh minat program memiliki kepribadian yang sama yaitu kepribadian conscientiousness dan peserta kurang dalam kepribadian Openess yang berarti peserta TIDAK memiliki sifat yang kreatif, Imajinatif, Intelektual, penasaran dan berpikiran luas. Selain itu juga kebanyakan minat untuk mengikuti program di MES ini kebanyakan berusia muda. Dari sini perusahaan bisa menawarkan program-program tertentu khususnya untuk usia 20-22 yang memiliki minat yang tinggi untuk mengikuti program-program yang ada di MyEduSolve



Suggestion

Dari hasil analisis yang kita lakukan, diperlukan :

- sosialisasi untuk program-program yang minatnya sedikit supaya lebih dikenal masyarakat
- Sosialisasi ke daerah daerah diseluruh Indonesia khususnya luar pulau jawa supaya program di MES lebih dikenal masyarakat
- Penawaran program serta sosialisasi kepada remaja yang baru lulus SMA serta Maba untuk mengikuti program yang ada di MES
- Mentor dapat memberikan tugas untuk meningkatkan kreatifitas peserta yang dimana hasil analisis menunjukan bahwa peserta kurang memiliki kepribadian Openess yang berarti peserta kurang memiliki sifat yang kreatif, Imajinatif, Intelektual, penasaran dan berpikiran luas.





Have a
great day
ahead.