LẬP TRÌNH PHÂN TÍCH DỮ LIỆU

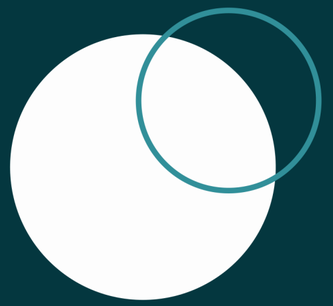
Data: Điểm thi THPT Quốc gia năm 2021 và 2022 tỉnh Lạng Sơn

NHÓM 2:

Thành viên:
Phan Lê Hoàng Việt
Hoàng Đức Chiến
Lương Quang Khải
Lưu Hoàng Ngọc Trinh

Giảng viên: Trương Vĩnh Linh

Thuyết minh dữ liệu



- Nguồn dữ liệu vietnamnet.vn.
- Thời gian thu thập: tháng 11/2022, dữ liệu được thu thập dùng để phân tích phổ điểm giữa hai kì thi THPT quốc gia 2021 và 2022.
- Gồm các feature SBD của thí sinh, Thí sinh thuộc Sở GD nào, điểm các môn thi của từng thí sinh (Toán, Văn, Sử, Địa, Lý, Hóa, Sinh, Ngoại Ngữ, GDCD) Và cuối cùng là label năm thi.

Xử lý dữ liệu

- Bộ dữ liệu có các cột không có giá trị do thí sinh chỉ đăng ký thi 3 môn bắt buộc (Toán, Văn, Ngoại ngữ) và khối Tự Nhiên (Lý, Hóa, Sinh) hoặc khối Xã Hội (Sử, Địa, Công dân).

- Ngoài ra còn có các thí sinh tự do chỉ đăng ký tối thiểu 3 môn để xét điểm đại học. Để tiện tính toán, ta tạm thời thay các giá trị trống bằng giá trị 0 và chia bộ dữ liệu thành 2 khối Tự Nhiên và Xã Hội. Các thí sinh THPT là các thí sinh phải thi đủ 6 môn của khối Tự Nhiên hoặc khối Xã Hội

Bộ dữ liệu sau khi xử lý

	SBD	So GD&DT	Toan	Van	Su	Dia	Ly	Hoa	Sinh	NgoaiNgu	GDCD	NamThi
0	10009389	So GD&DT tinh Lang Son	2.8	2.75	3.00	6.75	0.0	0.0	0.0	3.6	6.00	2022
1	10009390	So GD&DT tinh Lang Son	3.0	4.25	5.00	4.00	0.0	0.0	0.0	0.0	0.00	2022
2	10009391	So GD&DT tinh Lang Son	6.4	7.00	6.00	7.50	0.0	0.0	0.0	2.6	8.50	2022
3	10009392	So GD&DT tinh Lang Son	4.6	5.50	7.00	6.25	0.0	0.0	0.0	0.0	0.00	2022
4	10009393	So GD&DT tinh Lang Son	7.0	6.00	7.25	7.00	0.0	0.0	0.0	3.8	8.50	2022
5	10009394	So GD&DT tinh Lang Son	3.8	5.75	5.00	6.25	0.0	0.0	0.0	2.4	6.00	2022
6	10009395	So GD&DT tinh Lang Son	4.0	3.25	4.75	7.25	0.0	0.0	0.0	0.0	0.00	2022
7	10009396	So GD&DT tinh Lang Son	5.4	6.00	6.00	7.50	0.0	0.0	0.0	0.0	0.00	2022
8	10009397	So GD&DT tinh Lang Son	6.8	6.75	7.25	7.00	0.0	0.0	0.0	3.0	8.25	2022
9	10009398	So GD&DT tinh Lang Son	2.6	3.50	4.50	5.25	0.0	0.0	0.0	0.0	0.00	2022
10	10009399	So GD&DT tinh Lang Son	6.2	6.00	6.25	7.50	0.0	0.0	0.0	2.4	7.50	2022
11	10009400	So GD&DT tinh Lang Son	5.4	5.00	5.00	6.50	0.0	0.0	0.0	0.0	0.00	2022
12	10009401	So GD&DT tinh Lang Son	3.8	4.50	4.25	6.25	0.0	0.0	0.0	0.0	0.00	2022
13	10009402	So GD&DT tinh Lang Son	3.4	6.25	6.75	8.25	0.0	0.0	0.0	0.0	0.00	2022
14	10009403	So GD&DT tinh Lang Son	4.2	4.50	3.75	5.75	0.0	0.0	0.0	2.0	5.50	2022

THỐNG KÊ SƠ BỘ

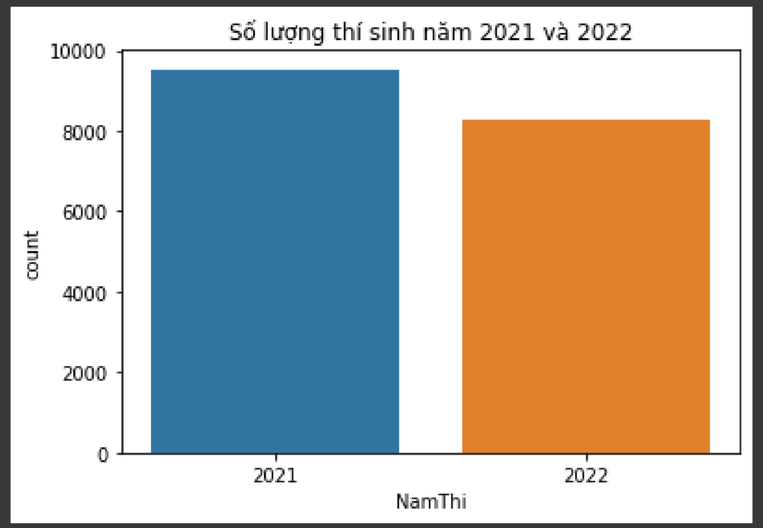
Đưa ra số lượng thí sinh
năm 2021 và 2022 và các phổ điểm

1) Số lượng thí sinh năm 2021 và năm 2022

```
[ ]   datanamthi = data_new[['NamThi']]
      print('Thí sinh năm 2021: ', datanamthi[datanamthi['NamThi'] == 2021].count())
      print('Thí sinh năm 2022: ', datanamthi[datanamthi['NamThi'] == 2022].count())
```

```
Thí sinh năm 2021:  NamThi    9530
dtype: int64
Thí sinh năm 2022:  NamThi    8284
dtype: int64
```

```
[ ]   sns.countplot(x='NamThi', data=data_new)
      plt.title('Số lượng thí sinh năm 2021 và 2022')
      plt.show()
```



Thống kê điểm

	Toan	Van	NgoaiNgu	Ly	Hoa	Sinh	NamThi
279	8.6	5.75	6.6	6.75	2.5	5.25	2021
280	9.4	7.50	9.6	7.50	6.5	5.50	2021
282	8.8	6.50	9.8	8.25	2.5	5.75	2021
283	9.2	5.25	8.8	5.00	9.5	9.75	2021
289	8.6	5.25	7.8	8.50	3.5	4.75	2021
292	8.8	6.50	9.4	8.25	7.5	4.50	2021
294	6.6	4.25	9.8	5.75	2.5	5.25	2021
296	7.6	7.00	9.4	8.50	2.5	3.25	2021
298	8.8	5.50	9.4	6.75	4.5	4.50	2021
301	9.6	7.25	9.0	8.00	7.5	5.50	2021

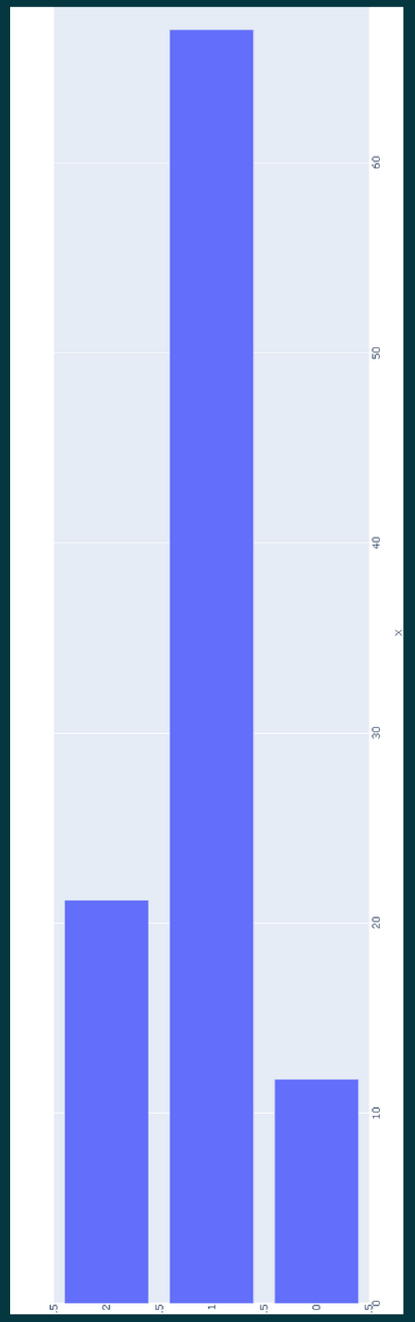
	Toan	Van	NgoaiNgu	Ly	Hoa	Sinh	NamThi
45	7.0	6.25	5.8	7.50	1.5	4.25	2022
46	8.4	6.50	4.6	7.50	5.5	4.50	2022
77	8.0	5.00	5.0	6.50	8.5	5.75	2022
81	6.0	3.50	6.4	5.50	2.5	3.25	2022
95	7.6	8.00	4.2	6.50	8.5	5.50	2022
98	8.2	7.00	4.0	7.50	7.5	5.00	2022
135	7.6	6.25	4.2	7.50	4.5	6.75	2022
152	7.4	6.25	4.4	7.00	2.5	7.50	2022
218	8.2	7.75	5.2	8.25	3.5	6.25	2022
231	8.6	4.00	4.0	7.75	5.5	3.75	2022

DASHBOARD

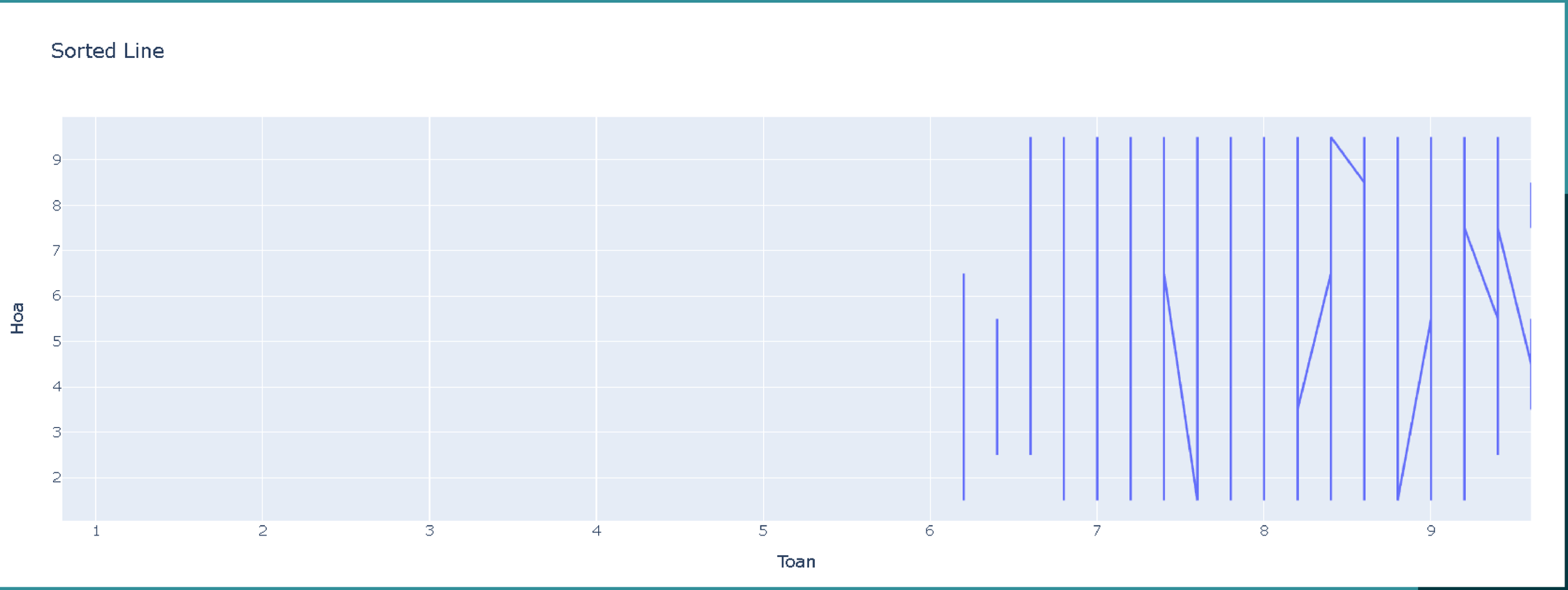
```
khoi = ['Khối Tự Nhiên','Khối Xã Hội', 'Thí Sinh tự do']
thisinh = [11.79,67,21.21]

fig = px.bar(khoi, thisinh)

fig.show()
```



```
df = data.sort_values(by="Toan")
fig = px.line(df, x="Toan", y="Hoa", title="Sorted Line")
fig.show()
```



Select box

```
mc=pw.Pywedge_Charts(data, c=None,y='NgoaiNgu')
chart= mc.make_charts()
```