

```
In [3]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import pprint

pr = pprint.PrettyPrinter()

# pd.set_option('display.max_columns', None)
# pd.set_option('display.max_rows', 5)

data = pd.read_csv('salary_.csv', encoding="cp1251", sep=';')

data.columns = ['Tabnum', 'Position', 'Job_rate', 'Dep_code', 'Main_sal', 'Add_sal']

# pr.pprint(data)
# pr.pprint(data.info())
```

```
In [4]: print("Фонд оплаты труда: ", data['Total_sal'].sum())
print()
print("Фонд оплаты труда по должностям:")

df = data.groupby('Position')['Total_sal']
print(df.sum())
print()

print('Фонд оплаты труда без внешних совместителей:')
df1 = data.groupby('Position')['Main_sal']
print(df1.sum())
print()

print('Фонд оплаты труда только для внешних совместителей:')
df2 = data.groupby('Position')['Add_sal']
print(df2.sum())
print()

print("Структурный состав персонала по должностям:")
print(data.Position.value_counts())
print("Общее количество:      ", data.Position.value_counts().sum())
print()
```

Фонд оплаты труда: 12018196.17

Фонд оплаты труда по должностям:

Position	
Ассистент	878028.88
Доцент	6227156.38
Преподаватель	564664.90
Профессор	1540199.77
Старший преподаватель	2808146.24

Name: Total\_sal, dtype: float64

Фонд оплаты труда без внешних совместителей:

Position	
Ассистент	454854.16
Доцент	2153699.12
Преподаватель	255962.73
Профессор	367487.62
Старший преподаватель	1229265.96

Name: Main\_sal, dtype: float64

Фонд оплаты труда только для внешних совместителей:

Position	
Ассистент	142335.23
Доцент	930485.12
Преподаватель	66856.17
Профессор	222734.21
Старший преподаватель	543734.76

Name: Add\_sal, dtype: float64

Структурный состав персонала по должностям:

Position	
Доцент	337
Старший преподаватель	237
Ассистент	105
Профессор	80
Преподаватель	53

Name: count, dtype: int64

Общее количество: 812

```
In [6]: # Круговая диаграмма
vals = data.Position.value_counts()

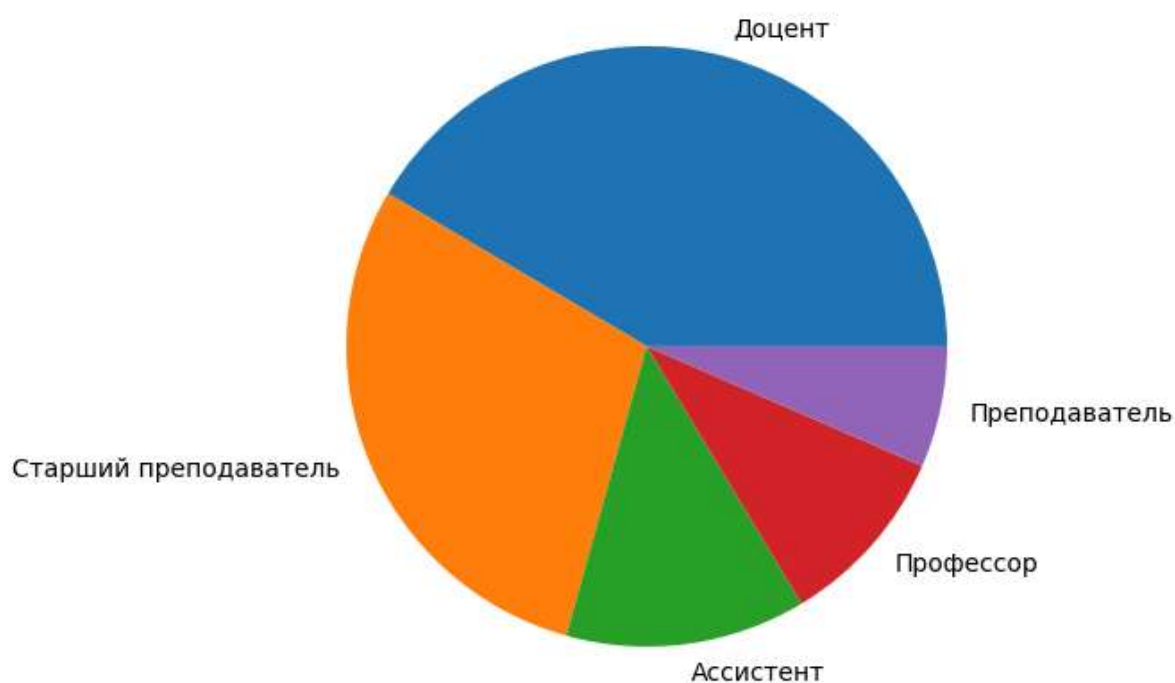
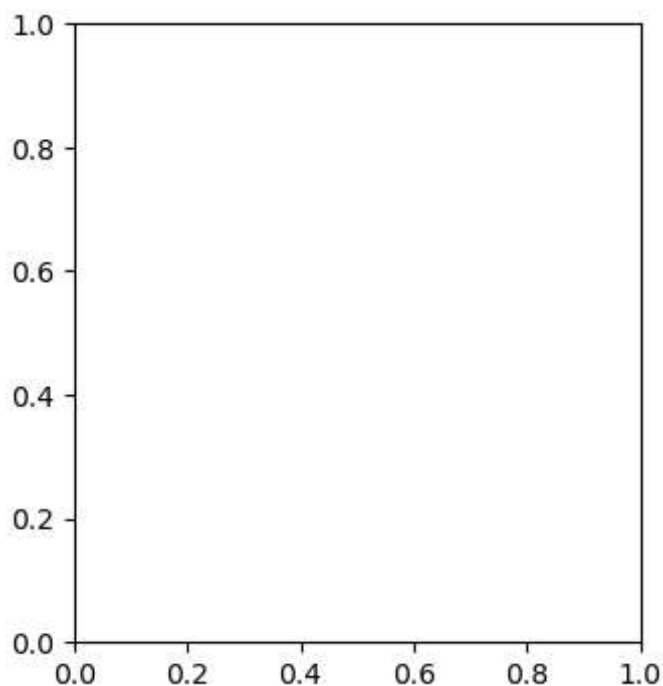
labels = ['Доцент', 'Старший преподаватель', 'Ассистент', 'Профессор', 'Преподаватель']
#labels = data.Position.unique() - этот метод выводит список уникальных имен не в том порядке

fig = plt.figure(figsize=(8,4))
plt.subplot(1, 2, 1)

fig, ax = plt.subplots()
ax.pie(vals, labels=labels)
ax.axis("equal")

ass = data[data['Position']=='Ассистент'].groupby('Tabnum')['Total_sal'].sum()
print("Мин.зарплата ассистента: ", ass.min())      # ???
print("Ср. зарплата ассистента: ", ass.mean())
print("Макс.зарплата ассистента: ", ass.max())
print()
```

Мин.зарплата ассистента: 334.39999999999964  
 Ср. зарплата ассистента: 8693.355247524752  
 Макс.зарплата ассистента: 21484.59



```
In [7]: # Количество доцентов, которые являются внешними совместителями (73)_
dotc1 = data[data['Position']=='Доцент']
dotc2 = dotc1[dotc1['Main_sal'].isna()]
print("Количество внешних доцентов-совместителей: ", dotc2['Add_sal'].count())

df = data[
    (data['Position']=='Доцент') &
    (data['Job_rate']) &
    (data['Main_sal'].isna())
].groupby('Tabnum')['Total_sal'].sum()

print(df)
# print(df['Add_sal'].count())    выводит тоже 73

print()
```

Количество внешних доцентов-совместителей: 73

Tabnum

527	4177.76
1032	1769.40
1053	2506.66
1244	6684.40
1250	3437.99

...

303395	21680.56
303459	12957.16
303483	3828.55
303495	8194.47
303537	3111.52

Name: Total\_sal, Length: 66, dtype: float64

```
In [8]: #5 Средняя зарплата внутреннего совместителя
df = data[data['Main_sal'].isna() != True].groupby('Tabnum')['Total_sal'].sum()
print("Мин.зарплата вн.совместителя: ", df.min())
print("Средн.зарплата вн.совместителя: ", df.mean())
print("макс. зарплата вн.совместителя:", df.max())

print()
```

Мин.зарплата вн.совместителя: 805.5

Средн.зарплата вн.совместителя: 17151.235353535354

макс. зарплата вн.совместителя: 45940.18

```
In [9]: #6 Средняя зарплата внешнего совместителя

df = data[data['Main_sal'].isna()].groupby('Tabnum')['Total_sal'].sum()
print("Мин.зарплата внешнего совместителя: ", df.min())
print("Средн.зарплата внешнего совместителя: ", df.mean())
print("макс. зарплата внешнего совместителя:", df.max())

print()
```

Мин.зарплата внешнего совместителя: 334.399999999999964

Средн.зарплата внешнего совместителя: 8972.36455882353

макс. зарплата внешнего совместителя: 50470.83

```
In [10]: #7 Найти доцентов, которые работают на 1,0 ст.
df = data[data['Position']=='Доцент'].groupby('Tabnum')['Job_rate'].sum()
print("Доценты, которые работают на 1,0 ст:\n", df[df==1.0])

print()
```

Доценты, которые работают на 1,0 ст:

Tabnum

```
114      1.0
540      1.0
585      1.0
1204     1.0
1504     1.0
1616     1.0
1651     1.0
2263     1.0
2293     1.0
2328     1.0
2383     1.0
2454     1.0
2774     1.0
2937     1.0
4544     1.0
5230     1.0
5703     1.0
70766    1.0
70812    1.0
74332    1.0
74635    1.0
74763    1.0
100364   1.0
106038   1.0
161208   1.0
164775   1.0
165021   1.0
165283   1.0
303567   1.0
```

Name: Job\_rate, dtype: float64

```
In [11]: #8 Найти самого незагруженного по основному месту работы
df = data[data['Main_sal'].isna() != True]

print("Самый незагруженный по основному месту работы:\n", df[df['Job_rate']==df['Job_rate']])
```

Самый незагруженный по основному месту работы:

	Tabnum	Position	Job_rate	Dep_code	Main_sal	Add_sal	Total_sal
308	303612	Ассистент	0.05	17054.0	400.95	NaN	805.5

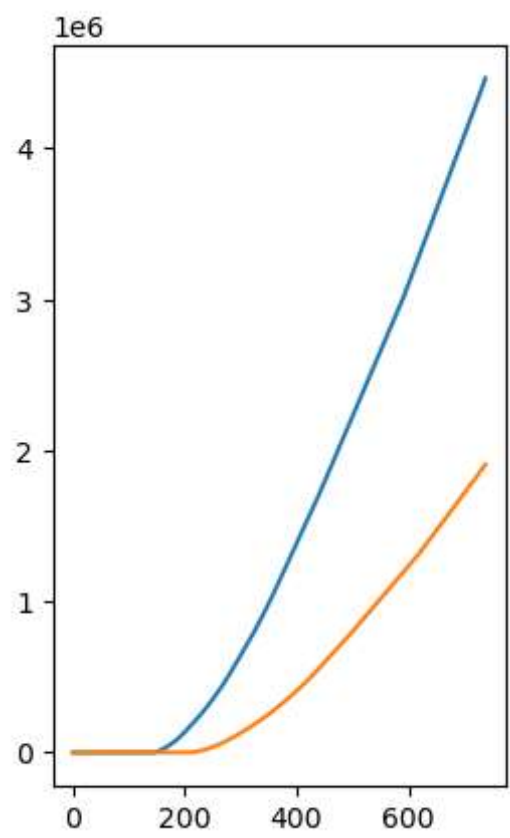
```
In [12]: #9 Нарисовать график Формирования ФОТ от меньших ЗП к большим
df = data.groupby('Tabnum')['Main_sal'].sum()
df = df.sort_values(ascending=True)
x = [i for i in range(len(df))]

# # df1 = data.groupby('Tabnum')['vn_sovm'].sum()
# # df1 = df1.sort_values(ascending=True)
# # x1 = [i for i in range(len(df1))]

df2 = data.groupby('Tabnum')['Add_sal'].sum()
df2 = df2.sort_values(ascending=True)
x2 = [i for i in range(len(df2))]

plt.subplot(1, 2, 2)
plt.plot(x, df.cumsum())
# plt.plot(x1, df1.cumsum())
plt.plot(x2, df2.cumsum())
```

Out[12]: [



```
In [ ]:
```