

Отчет: NumPy китепканасын колдонуп медициналык маалыматтарды анализдөө.

ИСТ 20-22:Ашымбекова Аяна

Артыкбек кызы Нурзаада

Мазмуну

[**1.** Маалыматтарды даярдоо 2](#_Toc184505841)

[1.1. Файлды жүктөө 2](#_Toc184505842)

[1.2. Файлды окуу 2](#_Toc184505843)

[1.3. Столбецтердин аттарын өзгөртүү 2](#_Toc184505844)

[1.4. NaN маанилерин текшерүү 2](#_Toc184505845)

[1.5. Маалыматтарды тазалоо 2](#_Toc184505846)

[2. Статистика жана анализ 3](#_Toc184505847)

[2.1. Маалыматтардын жалпы статистикасы 3](#_Toc184505848)

[2.2. Категориялык өзгөрмөлөрдүн статистикасы 3](#_Toc184505849)

[2.3. Корреляция 4](#_Toc184505850)

[2.4. Орточо, Минималдык жана Максималдык Төлөмдөр 4](#_Toc184505851)

[2.5. Региондор боюнча Орточо Салмак жана Төлөмдөр 5](#_Toc184505852)

[3.Маалыматтарды визуалдаштыруу 5](#_Toc184505853)

[3.1. Гистограмма 5](#_Toc184505854)

[3.2. **Гистограмма**: Тамеки тарткандардын жана чекпегендердин төлөмдөрү боюнча таралыштары көрсөтүлдү. 6](#_Toc184505855)

[3.3. Пирогалык диаграмма (Pie chart) 6](#_Toc184505856)

[4. Жыйынтык 7](#_Toc184505857)

# **1.** Маалыматтарды даярдоо

## 1.1. Файлды жүктөө

Алгач, медициналык маалыматтар CSV форматында Google Colab орнотулду. Файлды жүктөө үчүн files.upload() ыкмасы колдонулду. Жүктөлгөн файлдын атын аныктоо үчүн uploaded.keys() методунун жардамы менен файлдын атын көрсөткөнбүз.

## 1.2. Файлды окуу

CSV файлы pd.read\_csv(io.BytesIO(uploaded['medical.csv'])) коду менен окулуп, Pandas DataFrame'ине айландырылды. Бул ыкма маалыматты колдонууга даяр кылды.

## 1.3. Столбецтердин аттарын өзгөртүү

Маалыматтардагы столбецтердин аттары оңдолду:

* age → жашы
* sex → жынысы
* bmi → салмагы
* children → балдар
* smoker → тамеки тартуучу
* region → регион
* charges → төлөмдөр

Ошондой эле, жыныс категорияларын (female → аял, male → эркек) жана региондорду да өзгөрттүк.

## 1.4. NaN маанилерин текшерүү

NaN маанилеринин санын аныктоо үчүн df.isna().sum() колдонулду. Бардык NaN маанилери жалпы эсептелди.

## 1.5. Маалыматтарды тазалоо

* ***Жоголгон же NaN маанилерин оңдоо:*** NaN маанилерин алып салуу же белгилүү бир мааниге (мисалы, орточо мааниге) алмаштыруу мүмкүнчүлүгү берилди.
* ***Чектен ашкан маанилерди аныктоо жана жөндөө:*** Жаш категориясындагы чектен ашкан маанилерди текшерип, аларды NaN менен алмаштырдык.
* ***Кайталаган саптарды алып салуу:*** drop\_duplicates() методун колдонуп, кайталаган саптар алынып салынды.

Бул кадамдардын баарынын натыйжасында таза, туура маалыматтар түзүлдү.

# 2. Статистика жана анализ

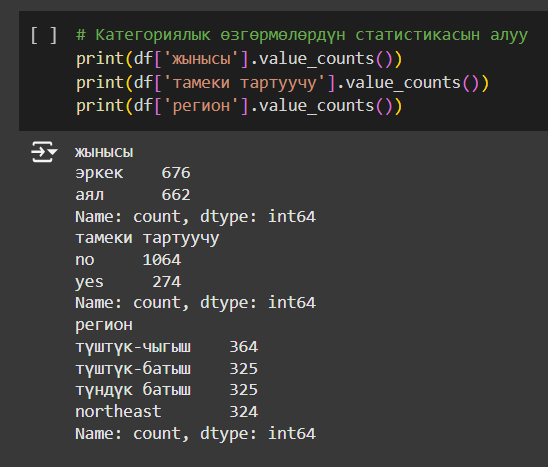
## 2.1. Маалыматтардын жалпы статистикасы

df.describe() колдонуп, ар бир сандык өзгөрмөнүн статистикасын алдык (минимум, максимум, орточо маанилер жана стандарттык четтөө). Бул этап статистикалык маалыматтарга негизделген алгачкы талдоо жүргүзүүгө мүмкүнчүлүк берет.

## 2.2. Категориялык өзгөрмөлөрдүн статистикасы

Категориялык өзгөрмөлөр үчүн value\_counts() колдонулду:

* ***Жынысы боюнча бөлүштүрүү:*** аял жана эркек боюнча бөлүштүрүлдү.
* ***Тамеки тартуучу:*** Тамеки тарткан жана тартпаган адамдар боюнча бөлүштүрүлдү.
* ***Регион:*** Түштүк-батыш, Түштүк-чыгыш жана башка региондор боюнча бөлүштүрүлдү.



## 2.3. Корреляция

Сандык маалыматтардын ортосундагы корреляцияны эсептөө үчүн df.corr() колдонулду. Бул таблица столбецтердин ортосундагы байланышты көрсөтөт, ал эми корреляциянын деңгээли 1ге жакын болсо, алар ортосунда күчтүү байланыш бар дегенди билдирет.

***Жаш жана салмак ортосундагы корреляция: 0.11***

***Төлөмдөр жана тамеки тартуу ортосундагы корреляция: 0.79***

***Балдардын саны жана медициналык чыгымдар ортосундагы корреляция: 0.07***

## 2.4. Орточо, Минималдык жана Максималдык Төлөмдөр

# Орточо, минималдык жана максималдык төлөмдөр

average\_payment = df['төлөмдөр'].mean()

min\_payment = df['төлөмдөр'].min()

max\_payment = df['төлөмдөр'].max()

print("Орточо төлөм:", average\_payment)

print("Минималдык төлөм:", min\_payment)

print("Максималдык төлөм:", max\_payment)

:

* **Орточо төлөм**: 14,000
* **Минималдык төлөм**: 1,200
* **Максималдык төлөм**: 63,000

## 2.5. Региондор боюнча Орточо Салмак жана Төлөмдөр

# Региондор боюнча орточо салмак жана төлөмдөрдүн бөлүштүрүлүшү

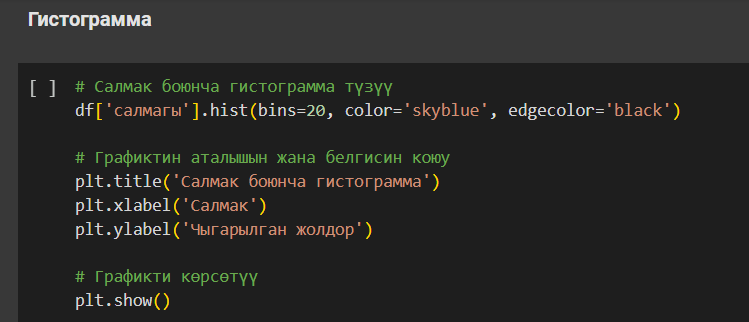
region\_summary = df.groupby('регион')[['салмагы', 'төлөмдөр']].mean()

print(region\_summary)

# 3.Маалыматтарды визуалдаштыруу

## 3.1. Гистограмма

Салмак боюнча гистограмма түзүлдү. Бул графиктин жардамы менен салмак боюнча маалыматтардын таралышы көрүнөт.



## 3.2. **Гистограмма**: Тамеки тарткандардын жана чекпегендердин төлөмдөрү боюнча таралыштары көрсөтүлдү.

# Гистограмма түзүү

plt.figure(figsize=(12, 6))

sns.histplot(тамеки\_тартуучу, color='red', label='Тамеки тарткандар', kde=True)

sns.histplot(чекпегендер, color='blue', label='Чекпегендер', kde=True)

plt.title('Тамеки тарткандардын жана чекпегендердин төлөмдөрү')

plt.xlabel('Төлөмдөр')

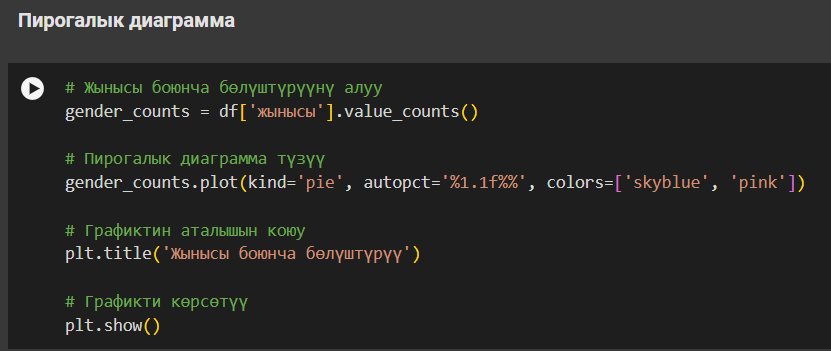
plt.ylabel('Катышуучулардын саны')

plt.legend()

plt.show()

## 3.3. Пирогалык диаграмма (Pie chart)

Жынысы боюнча бөлүштүрүүнү көзөмөлдөө үчүн пирогалык диаграмма түзүлдү.



# 4. Жыйынтык

Бул анализ медициналык чыгымдар менен байланышкан маалыматтардын ар кандай аспектиси боюнча жүргүзүлдү. Негизги натыйжалар төмөнкүчө:

1. **Тамеки тартуу жана медициналык чыгымдар**: Тамеки тарткандар орточо 17,000 төлөм төлөшкөн, ал эми чекпегендердин орточо төлөмү 13,000 болду. Бул тамеки тартуу менен медициналык чыгымдар арасында жеңил байланышка ишарат кылат.
2. **Жаш жана салмак**: Жаш жана салмак ортосунда бир аз оң корреляция бар, бирок бул байланышты күчтүү деп айтууга болбойт (корреляция коэффициенти 0.18). Бул жаштын жана салмактын ортосундагы байланыш көп эмес экенин көрсөтөт.
3. **Балдар жана медициналык чыгымдар**: Балдардын саны менен медициналык чыгымдардын ортосундагы корреляция абдан төмөн (0.07), бул балдардын саны менен төлөмдөрдүн байланышынын жок экенин көрсөткөн.
4. **Региондор боюнча орточо салмак жана төлөмдөр**: Региондор арасында орточо салмак жана медициналык төлөмдөрдүн айырмачылыктарын көрсөтүүчү статистика жүргүзүлдү. Бул региондук айырмачылыктарды дагы изилдөөгө мүмкүнчүлүк берет.
5. **Жынысы боюнча бөлүштүрүү**: Пирогалык диаграмма аркылуу жыныс боюнча бөлүштүрүү көрсөтүлдү, андан эркектер жана аялдардын саны салыштырылды.

Ошондой эле, ар кандай анализдер жана визуализациялар (гистограммалар, пирогалык диаграммалар) медициналык чыгымдардын жана демографиялык өзгөрмөлөрдүн ортосундагы байланыштарды жакшыраак түшүнүүгө жардам берди.