ДЕРЕКТЕРДІ ИНТЕЛЛЕКТУАЛДЫ ТАЛДАУ

Дәріс 8

pandas және Matplotlib арқылы визуализация

Оспан Ә.Ғ., PhD, доцент-зерттеуші

ҚазҰУ, Жасанды интеллект және Big Data кафедрасы

2026-2027 оқу жылы

Дәрістің мақсаты мен мазмұны

Мақсат және оқу нәтижелері

- Сұрақ пен белгі типтеріне қарай график түрін таңдау.
- Графикті безендіруге дейін агрегаттауды тексеру.
- Жазулары мен шектеулері бар, оқуға ыңғайлы суреттер жасау.

Дәріс мазмұны

1. **График таңдау**
2. **Figure және Axes**
3. **Үлестірімдер**
4. **Байланыстар мен топтар**
5. **Композиция және экспорт**

Аналитикалық сұрақ бойынша график таңдау

Сұрақ	Дайындық	График
Санатта қанша жазба бар?	value_counts	Бағандар
Сан қалай үлестірілген?	Сандық баған	Гистограмма
Топтар қалай ерекшеленеді?	Топтар бойынша мәндер	Boxplot
Екі сан қалай байланысты?	Бірдей жолдардағы жұптар	Scatter
Уақыт ішінде не өзгереді?	Күндерді тану және сұрыптау	Сызық
Бірнеше кесіндіні қалай салыстыруға болады?	Бірдей ережелер	Панельдер торы

Figure және Axes: басқарылатын сурет

Figure

- Тұтас сурет және оның жалпы өлшемі.
- Бір немесе бірнеше аймақты қамтиды.
- fig.savefig нақты нәтижені сақтайды.

Axes

- Деректері бар бір координаттар жүйесі.
- ax.set жазулар мен тақырыпты береді.
- ax.legend топтардың шартты белгілерін түсіндіреді.

fig, ax = plt.subplots() сурет пен аймақты айқын байланыстырады.

Позициялар жиілігі: алдымен кесте

Python мысалы

```
from pathlib import Path
import pandas as pd
import matplotlib.pyplot as plt

football = pd.read_csv(Path('data') / 'football.csv')
counts = football['Position'].value_counts().head(8)
print(counts)
fig, ax = plt.subplots(figsize=(8, 4), layout='constrained')
counts.plot.bar(ax=ax, color='steelblue')
ax.set(title='Ең жиі кездесетін 8 позиция',
       xlabel='Позиция', ylabel='Жазбалар саны')
plt.show()
```

Дайындық

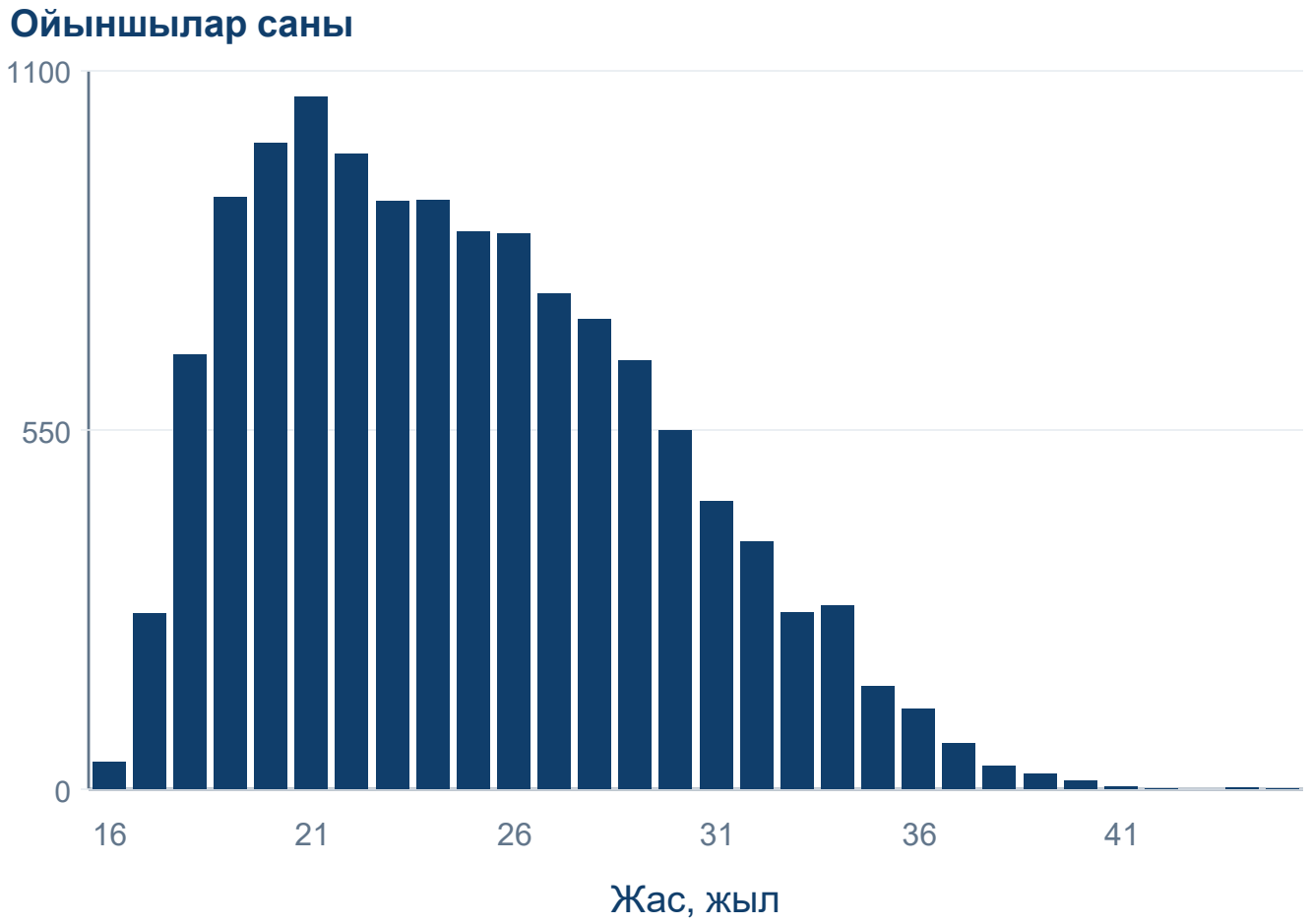
- Жиіліктер график салуға дейін есептелген.
- Бүкіл құрам емес, сегіз санат көрсетілген.

Жазулар

- Белгі және санауыш бірлігі көрсетілген.
- Кесте баған биіктіктерін тексеруге мүмкіндік береді.

Графикалық салыстыру нақты сандық дереккөзбен бірге беріледі.

Жас гистограммасы: пішіні нені көрсетеді



Үлестірімді қалай оқу керек

- Әр баған бір жас аралығын біріктіреді.
- Биіктік себептік механизмді емес, жазбалар санын көрсетеді.
- bins ені бөлшектердің көрінуіне әсер етеді.
- Жас құрамы тарихи кескінге қатысты.

Себептерді түсіндіруден бұрын байқалған пішінді тіркеп, оны жиілік кестесімен тексеріңіз.

Гистограмма, бағандар және тығыздық

Кескін	x осі	Биіктіктің мағынасы
Санаттар бағандары	Жеке атаулар	Жиілік немесе агрегат
counts гистограммасы	Сандық аралықтар	Бақылаулар саны
density гистограммасы	Сандық аралықтар	Тығыздық; аудан 1-ге тең
Топтарды салыстыру	Ортақ bins шектері	Нормалауға тәуелді

Пішінді салыстыру үшін көлемдерді ескеріп, бірдей аралықтарды қолданыңыз.

Аралық шектері: санауыштарды тексереміз

Python мысалы

```
import numpy as np

ages = np.array([18, 19, 21, 22, 23, 24])
edges = np.array([18, 21, 25])
hist_counts, _ = np.histogram(ages, bins=edges)
print(hist_counts.tolist(), hist_counts.sum())
fig, ax = plt.subplots(layout='constrained')
ax.hist(ages, bins=edges, edgecolor='black')
ax.set(xlabel='Жас, жыл', ylabel='Бақылаулар саны')
plt.show()
```

Нәтиже: [2, 4] 6

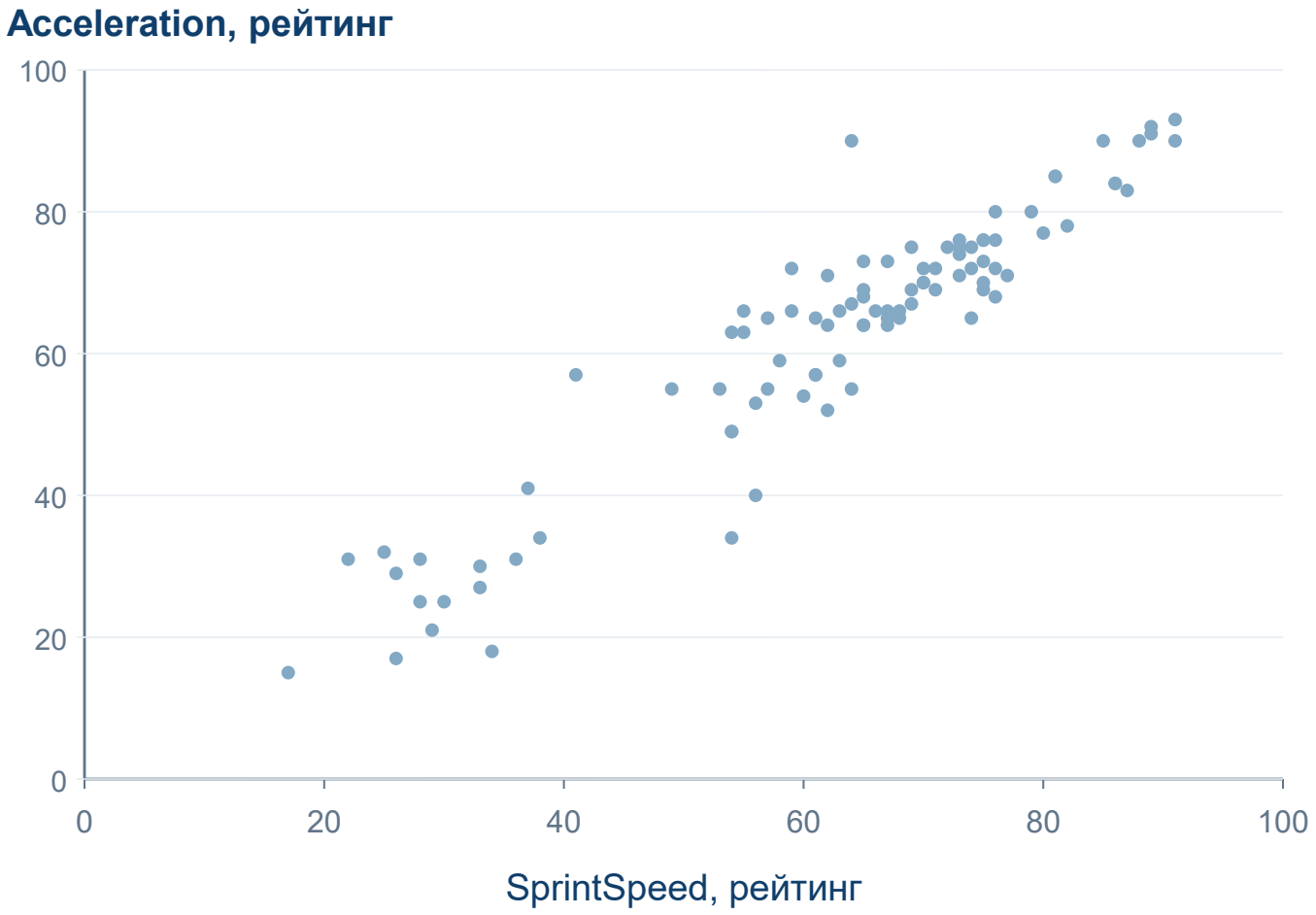
Шектер

- [18, 21) екі мәнді қамтиды.
- [21, 25] төрт мәнді қамтиды.

Бақылау

- Жиіліктер қосындысы көлемге тең.
- Шекаралық сан екі аралыққа түспейді.

Scatter: үдеу және жылдамдық



Нүкте нені білдіреді

- Координаттар бір жазбадан алынған.
- Мөлдірлік қабаттасуды байқауға көмектеседі.
- Рейтингтердің бірге өзгеруі себептілікті дәлелдемейді.
- Позиция мен бағалау жүйесі байланыстың бір бөлігін түсіндіруі мүмкін.

Ойыншыларды кездейсоқ жол ретімен сызықпен қоспаңыз: бұл рет уақытты білдірмейді.

Топтарды салыстыру: көрсеткіш және көлем

Нені салыстырамыз

- Орташа мәнді, медианаңы немесе жазбалар санын.
- Үлестер айқын бөлімді талап етеді.
- Ортақ шкалаға бірдей өлшем бірліктері қажет.

Қасында не хабарлаймыз

- Әр топтың n көлемін.
- Шашырау көрсеткішін және іріктеу ережесін.
- Таңдалған санаттардың толық жиынды қамтуын.

Топ көлемінсіз агрегатты шамадан тыс түсіндіру оңай.

Минимум мен максимум: екі баған қатары

Python мысалы

```
summary = football.groupby('Position')['Stamina'].agg(
    ['size', 'min', 'max']
)
shown = summary.loc[summary['size'].nlargest(6).index]
print(shown)
fig, ax = plt.subplots(figsize=(8, 4), layout='constrained')
shown[['min', 'max']].plot.bar(ax=ax)
ax.set(xlabel='Позиция', ylabel='Stamina, рейтинг',
       title='Жиі 6 позиция: ең аз және ең көп')
ax.legend(['Ең аз', 'Ең көп'])
plt.show()
```

Кесте

- Көлем мен экстремумдар айқын есептелген.
- Ең жиі алты позиция таңдалған.

Шектеу

- Құлаш әдеттегі деңгейді сипаттамайды.
- Ортаны сипаттауға медиана мен квартильдер қажет.

Boxplot: орта, қорап және ерекше нүктелер

Медиана және қорап

- Сызық: медиана.
- Қорап шеттері: Q1 және Q3; IQR = Q3 - Q1.

Мұрттар

- Шектер ішіндегі бақыланған мәндерге дейін.
- Әдеттегі шектер: Q1/Q3 және 1.5 IQR түзетуі.

Жеке нүктелер

- Мұрттардан тыс орналасады.
- Автоматты түрде қате болып саналмайды.

Boxplot үлестірімді толықтырады, бірақ оны толығымен алмастырмайды.

Панельдер торы және кірістірілген аймақ

subplots торы

- 2x3 алты аймақты білдіреді.
- axes[0, 1] жоғарғы ортаңғы панельді таңдайды.
- Ортақ шкалалар тікелей салыстыруды жеңілдетеді.

add_axes кірістірмесі

- [left, bottom, width, height].
- Барлық координат — Figure өлшемінің үлестері.
- Негізгі сурет пен жазуларды жаппайды.

Композиция бос орынды әшекейлемей, салыстыруды нақтылауы тиіс.

Экспорт: суретті есеп өлшемінде тексеру

1

Figure нысанын сақтау

Орынды өлшем мен dpi беріп, fig.savefig қолдану.

2

Нәтижені ашу

Жазуларды, түсіндірме белгілерін, қиылуды және мәтін өлшемін тексеру.

3

Сандарды салыстыру

Бағандар мен агрегаттарды бастапқы кестемен сәйкестендіру.

4

Түсіндірме қосу

Сұрақты, кесіндіні, өлшем бірліктерін және түсіндіру шектерін көрсету.

- Сурет деректері кескіндерден бөлек сақталады.
- Жергілікті нәтижелер work_results/ ішіне жазылады.

Визуализация қателері және олардың салдары

Қате	Бұрмалау	Түзету
Ойыншылар индексі бойынша сызық	Жалған тізбек	Scatter немесе үлестірім
Бағандардағы нөлдің қиылуы	Айырмашылықты үлкейту	Ұзындықты нөлден есептеу
Топтардағы әртүрлі bins	Пішіндер салыстырылмайды	Ортақ шектер
n көрсетілмеген орташа мән	Дерек көлемі жасырылған	Топ көлемдерін көрсету
Жазусыз ішінара кесінді	Толық қамту елесі	Ереже мен үлесті көрсету
Оқылмайтын экспорт	Нәтиже қолжетімсіз	Есеп өлшемінде қарау

Дәрістің негізгі нәтижелері

1. Сұрақ агрегаттауды және график түрін анықтайды.
2. Figure суретті қамтиды; Axes нақты координаттар жүйесін береді.
3. Гистограмма мағыналы аралықтар мен нормалауды талап етеді.
4. Топтарды салыстыру көлемдермен және шашыраумен бірге беріледі.
5. Экспорт көзбен және сандық кесте бойынша тексеріледі.

Бақылау сұрақтары және өзіндік жұмыс

1. Гистограмма санаттар бағандарынан қалай ерекшеленеді?
2. Футболшы жолының нөмірі неге уақыт емес?
3. density=True болғандағы биіктік нені білдіреді?
4. boxplot мұртынан тыс нүкте неге қатені дәлелдемейді?
5. Топтардың орташа мәндері жанында қандай мәліметтер қажет?

Өзіндік жұмыс

Әртүрлі bins арқылы Age үшін үш гистограмма құрыңыз. Тұрақты қасиеттерді және безендіруге тәуелді бөлшектерді атаңыз; жиіліктер қосындысын тексеріңіз.

Дереккөздер және құжаттама

Құжаттама

1. pandas: визуализация

https://pandas.pydata.org/docs/user_guide/visualization.html

2. Matplotlib: қысқаша нұсқаулық

https://matplotlib.org/stable/users/explain/quick_start.html

3. Matplotlib: boxplot

https://matplotlib.org/stable/api/_as_gen/matplotlib.axes.Axes.boxplot.html

Дәрістің негізі

- Берілген 8-дәрістің презентациясы
- Берілген 10-дәрістің деректерді визуализациялау жөніндегі презентациясы
- Берілген 8-зертханалық жұмыс
- Берілген 8-зертханалық жұмыстың жауаптары