Exercise Session 2

Social Computing 2025

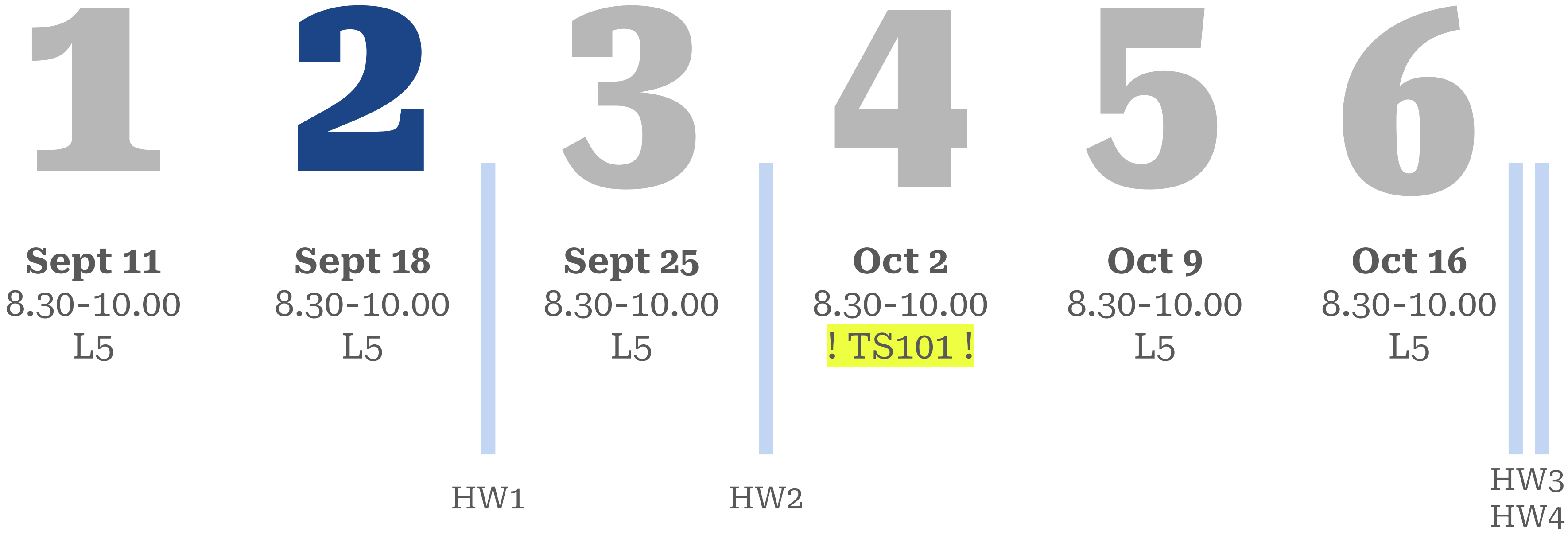
September 18, 8:30 - 10:00

Today’s Topics

1. Introduction
2. Schedule
3. Coursework Project
4. Homework 1
5. Q&A

1. Schedule

Exercise Schedule



1
Sept 11
8.30-10.00
L5

2
Sept 18
8.30-10.00
L5

HW1

3
Sept 25
8.30-10.00
L5

HW2

4
Oct 2
8.30-10.00
! TS101 !

5
Oct 9
8.30-10.00
L5

6
Oct 16
8.30-10.00
L5

HW3
HW4

See more on Moodle.

3. Coursework Project

Coursework Project Structure

- Consists of 4 homeworks, each with their own deadlines
- Must use coursework template
- Must work individually
- First HW due 22.9.2025 23:59
- Total 15 + 15 + 15 + 20 = 65 points (65% of course grade)
- Largely based on Mini Social

4. In-class Exercise

Reminder: restricted AI use in this course

AI must not be used for

- Ideation → Use your head
- Coding → Use documentation
- Writing → Do your best
- Debugging → Gain experience
- Course questions → Contact me

"I wish I didn't use ChatGPT in the Social Computing course. I thought I could get away with it."

– Anonymous Student, 2024
(currently in Oulu Prison)

Requirements

- I will be using Python 3.9.18. Any python is OK as long as it works.
- Using Jupyter notebooks is fine
- Disable AI features in IDE
- Python: https://docs.python.org/3.9/
- Sqlite3: https://docs.python.org/3/library/sqlite3.html
- Pandas: https://pandas.pydata.org/docs/
- SQL syntax: https://www.sqlite.org/lang.html
    - Help: Learning SQL: Master SQL Fundamentals by Alan Beaulieu
- Obtain database file from Moodle
- Feel free to use search engines, programming forums, tutorials etc.

SQL Fundamentals

- Help: Learning SQL: Master SQL Fundamentals by Alan Beaulieu
  - p. 4 – p. 8 → Introduction to SQL
  - p. 41 – p. 45 → SQL queries
- SQLite syntax: https://www.sqlite.org/lang.html
- Must know at least:
  - Basic operations on existing tables:

    SELECT, INSERT, DELETE, UPDATE
  - Clauses:

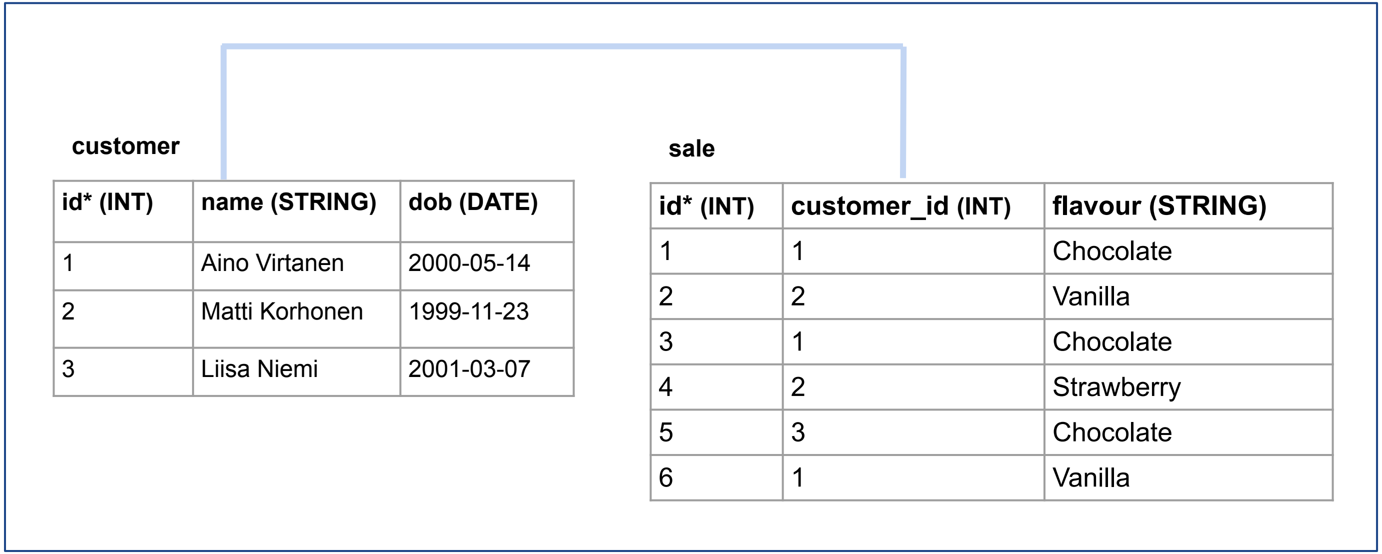
    ORDER BY, GROUP BY, LIMIT
  - Compound operations:

    UNION, JOIN

Note
You can still achieve full marks on the coursework by using only Python, but you should take this as an opportunity to practice SQL and put it on your CV.

SQL Fundamentals (refresher)

icecreamshop



customer

id* (INT)	name (STRING)	dob (DATE)
1	Aino Virtanen	2000-05-14
2	Matti Korhonen	1999-11-23
3	Liisa Niemi	2001-03-07

sale

id* (INT)	customer_id (INT)	flavour (STRING)
1	1	Chocolate
2	2	Vanilla
3	1	Chocolate
4	2	Strawberry
5	3	Chocolate
6	1	Vanilla

SQL Fundamentals (refresher)

```
SELECT * FROM customer;
```

result

id	name	dob
1	Aino	2000-05-14
2	Matti	1999-11-23
3	Liisa	2001-03-07

```
SELECT flavour, COUNT(*) AS times_bought
FROM sale
GROUP BY flavour;
```

result

flavour	times_bought
Chocolate	3
Vanilla	2
Strawberry	1

```
SELECT customer.name, COUNT(*) AS total_purchases
FROM sale
JOIN students
     ON sale.customer_id = customer.id
GROUP BY customer.id
ORDER BY total_purchases DESC
LIMIT 1;
```

result

name	total_purchases
Aino Virtanen	3

Pandas 1/2

pypi.org/project/pandas

1 – Install

```
pip install pandas
```

2 – Import

```
import pandas
```

3 – Create DF

```
data = {
     'name': ['Matti', 'Bob', 'Charlie', 'David'],
     'age': [25, 32, 28, 45],
     'city': ['Oulu', 'Paris', 'London', 'Tokyo']
}
df = pd.DataFrame(data)
```

4 – Inspect DF

```
df.info()
print(df.head(2))
print(df.tail(2))
```

Pandas 2/2

1 – Select columns

```
ages = df['age']
name_and_city = df[['name', 'city']]
```

2 – Filter

```
over_30 = df[df['age'] > 30]
```

3 – Update

```
df['age_next_year'] = df['age'] + 1
```

4 – Analyse

```
average_age = df['age'].mean()
```

Pandas vs SQL

Pandas

```python
name_and_salary = df[['Name', 'Salary']]
```

SQL

```sql
SELECT Name, Salary FROM employees;
```

Pandas

```python
total_employees = len(df)
unique_cities_count = df['City'].nunique()
```

SQL

```sql
SELECT
    COUNT(*) AS total_employees,
    COUNT(DISTINCT city) AS unique_cities
FROM employees;
```

Matplotlib 1/2

pypi.org/project/matplotlib

1 – Install

```
pip install matplotlib as plt
```

2 – Import

```
import matplotlib
# also pandas
```

3 – Prepare data

```
data = {
     'x_values': [0, 1, 2, 3, 4, 5],
     'y_values': [0, 1, 4, 9, 16, 25]
}
df = pd.DataFrame(data)
```

4 – Create plot

```
plt.plot(df['x_values'], df['y_values'])
```

5 – Display plot

```
plt.show()
```

Matplotlib 2/2

pypi.org/project/matplotlib

5 – Customise

```
plt.title('Y = X^2')
plt.xlabel('X')
plt.ylabel('Y')
plt.xticks(rotation=45)
plt.legend()
plt.grid(True)
```

6 – Other plot types

```
plt.scatter(df['x_values'], df['y_values'])
plt.bar(df['x_values'], df['y_values'])
```

Exercise 7

Task

“Lurkers”: Find the number of user pairs where user A left at least 5 reactions on user B’a posts but never left a comment. You may use SQL and/or Python to perform this task.

Time

20 minutes

Format

groups of 2 or 3

This exercise helps with HW tasks

1.4, 2.4

Exercise 8

Task

Find the minimum, average, median and maximum number of comments on posts. You may use SQL and/or Python to perform this task. Identify the post(s) with the most comments.

Time

15 minutes

Format

groups of 2 or 3

This exercise helps with HW tasks

1.3, 2.2, 2.3

Exercise 9

Task
You're the social media manager of @WildHorse. For each month since @WildHorse registered, calculate how comments they received total. Plot the cumulative values (time on X axis, number of comments on the Y axis). Based on the historical growth rate, predict how many months until the total number of comments on their posts reaches 200.

Time
30 minutes

Format
groups of 2 or 3

This exercise helps with HW tasks
2.1

5. Homework 1

Homework

- Finish all the in-class exercises (not graded)
- Download homework coursebook
- Task 1 – **due 22.9.2025 23:59**
- Worth 15 points. -1 point per day late penalty
- Submit on Moodle

7. Q&A, Customer Support