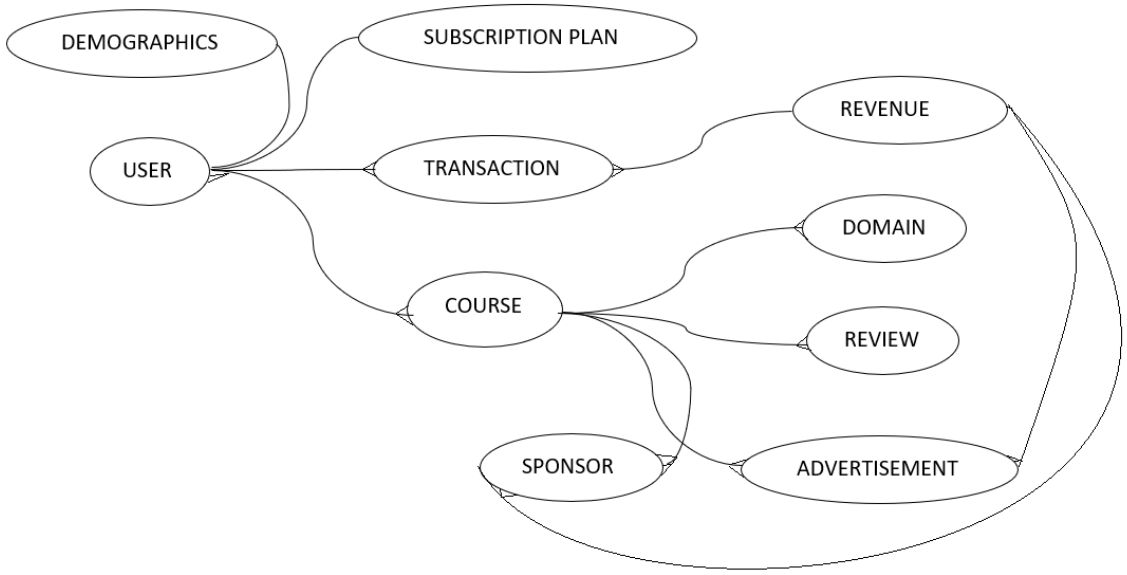
# Introduction to Data Analytics for Business

Final Course Assignment

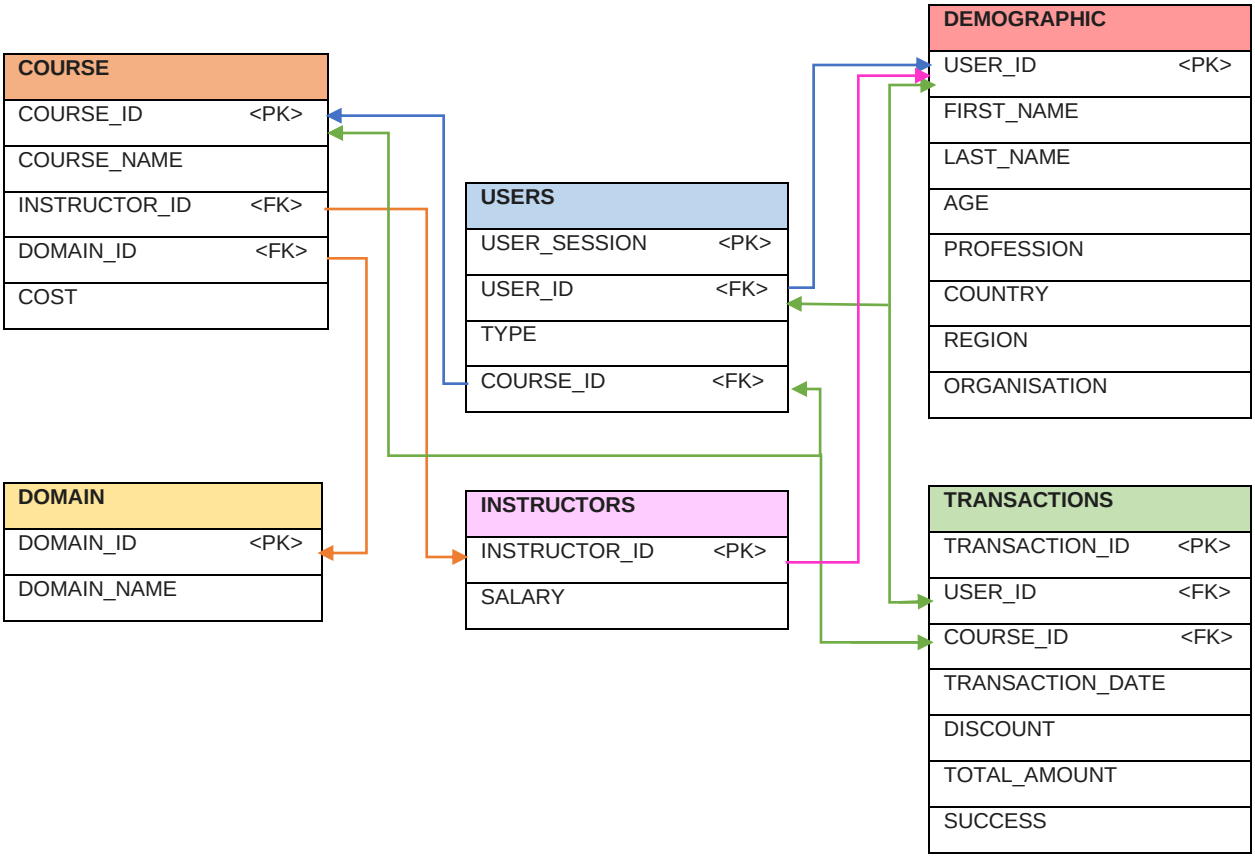
**Part 1: Conceptual business model.** Construct a conceptual business model for an industry or business that you are familiar with or have interest in.

**Business:** E-learning platform

1. **Instructors:** Can be students/teachers
2. **Learners:** Can be students/non-students
3. **Courses:** Domain-specific/Multiple Domains
4. **Subscription plans for learners:** Weekly/Monthly/Yearly/Biyearly
5. **Discounts on plans for instructors:** Weekly/Monthly/Yearly/Biyearly
6. **Non-Subscription plans for learners:** Per course
7. **Sponsorship for courses**
8. **Advertisement for courses:** Social Media/Email/Newsletters/etc.



**Part 2: Relational data model.** Take a subset of the ideas from the conceptual model you constructed in Part 1 and design a simple relationship model.



**Part 3: SQL queries.** Using the data model you constructed in Part 2, come up with two data extracts you think would be interesting, then write SQL queries to provide each one.

1. **NAME AND PROFESSION OF USERS WHO ARE BOTH LEARNERS AND INSTRUCTORS ON THE PLATFORM**

```
SELECT
    U.USER_ID,
    CONCAT (D.FIRST_NAME,' ',D.LAST_NAME) AS NAME,
    D.PROFESSION
FROM USERS U
JOIN DEMOGRAPHIC D
    ON D.USER_ID = U.USER_ID
GROUP BY 1
HAVING COUNT(DISTINCT TYPE) = 2
```

2. **MAXIMUM AND MINIMUM SALARY OF INSTRUCTORS IN EACH COUNTRY IN EACH DOMAIN**

```
SELECT
    DE.COUNTRY,
    DO.DOMAIN_NAME,
    MIN(I.SALARY) AS MINIMUM,
    MAX(I.SALARY) AS MAXIMUM
FROM COURSE C
JOIN DOMAIN DO
        ON C.DOMAIN_ID = DO.DOMAIN_ID
JOIN INSTRUCTORS I
    ON C.INSTRUCTOR_ID = I.INSTRUCTOR_ID
JOIN DEMOGRAPHIC DE
    ON C.INSTRUCTOR_ID = DE.USER_ID
GROUP BY DE.COUNTRY, DO.DOMAIN_ID, DO.DOMAIN_NAME
ORDER BY 1,2
```

3. **NAMES OF USERS WHO HAVE MADE TRANSACTIONS ATLEAST ONCE IN 15 DAYS**

```
SELECT
    T.USER_ID,
    CONCAT (D.FIRST_NAME,' ',D.LAST_NAME) AS NAME,
FROM TRANSACTIONS T
JOIN DEMOGRAPHIC D
    ON T.USER_ID = D.USER_ID
WHERE TRANSACTION_DATE >= DATE_SUB (CURDATE(), INTERVAL 15 DAY)
GROUP BY 1,2
```

**Part 4: Sensitive data and data quality issues.** Consider the data privacy and data quality implications of the data model you constructed in Part 2.

**1. DEMOGRAPHIC:** *AGE, PROFESSION, ORGANIZATION, COUNTRY, REGION* can be sensitive personal information

**2. DOMAIN:** *DOMAIN_NAME* can be ambiguous, unspecific, similar to other domains