DESAUTELS Faculty of Management
Faculté de gestion

# INDIVIDUAL PROJECT

**Due on September 10th, 2021**

**INSY 661: Database and Distributed Systems for Analytics**

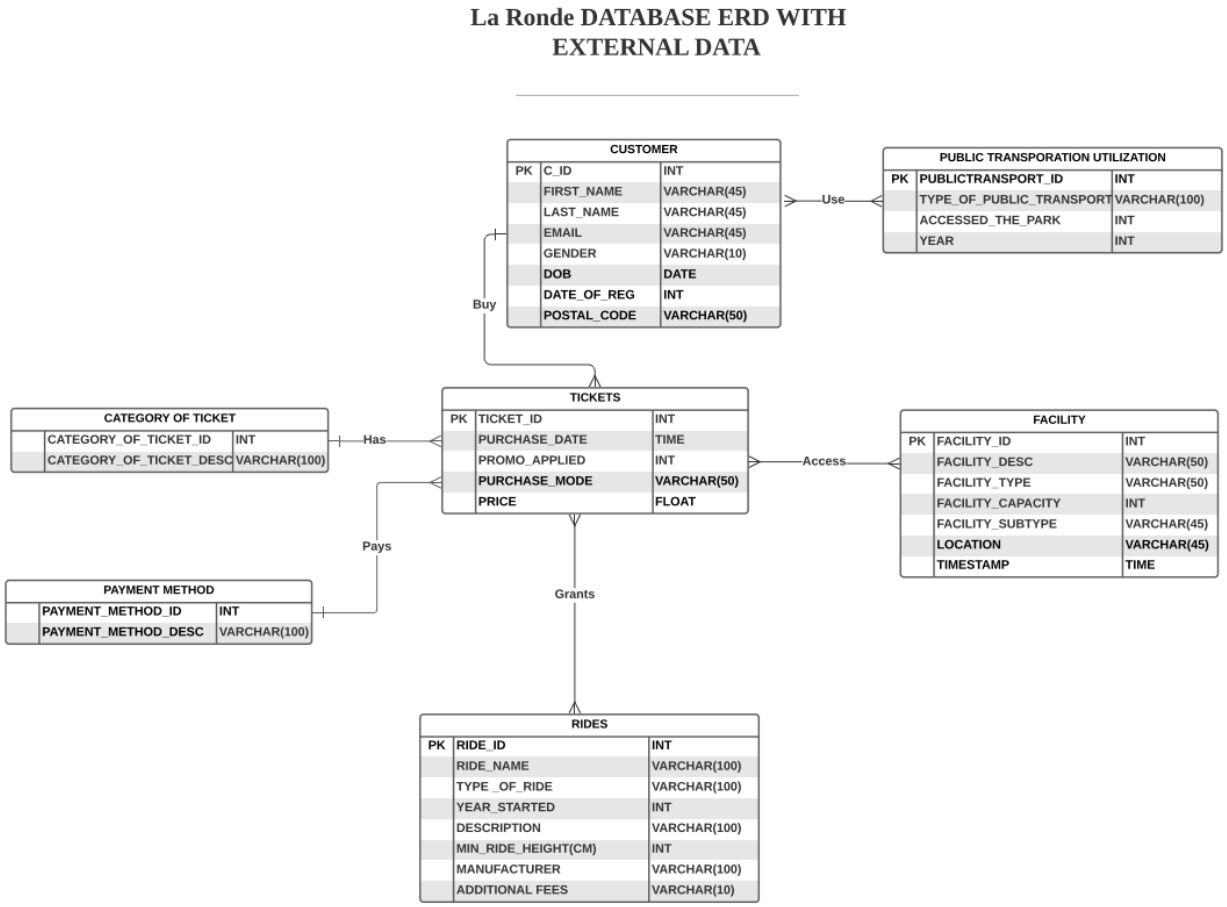
Professor Animesh Animesh

Summer 2021



| Student Name | Student ID |
|---|---|
| Mesaye Bahiru | 260634934 |

## SECTION-1: ERD with External data




**Assumptions**: Category of Ticket and Payment Method should be seen as a weak entity because they lacked unique primary keys. Tickets can be seen as a parent for these entities.

**External data (public transport utilization)**: public transportation usage to access the park

**SECTION-2: RELATIONAL MODEL**

**Foreign keys represented in ***

CUSTOMER (C_ID, FIRST_NAME, LAST_NAME, EMAIL, GENDER, DOB, DATE_OF_REG, POSTAL_CODE)
PK: C_ID
FK: N/A

PUBLIC TRANSPORATION UTILIZATION (PUBLICTRANSPORT_ID, TYPE_OF_PUBLIC_TRANSPORT, ACCESSED_THE_PARK, YEAR)

PK: PUBLICTRANSPORT_ID
FK: N/A

CUSTOMER_USE (C_ID*, PUBLICTRANSPORT_ID*)
PK: C_ID, PUBLICTRANSPORT_ID
FK: C_ID references CUSTOMER (C_ID)
PUBLICTRANSPORT_ID references PUBLIC TRANSPORATION UTILIZATION (PUBLICTRANSPORT_ID)

CATEGORY OF TICKET (CATEGORY_OF_TICKET_ID, CATEGORY_OF_TICKET_DESC)
PK: N/A
FK: N/A

PAYMENT METHOD (PAYMENT_METHOD_ID, PAYMENT_METHOD_DESC)
PK: N /A
FK: N /A

TICKETS (TICKET_ID, PURCHASE_DATE, PROMO_APPLIED, PURCHASE_MODE, PRICE, C_ID*, CATEGORY_OF_TICKET_ID*, PAYMENT_METHOD_ID*)

PK: TICKET_ID
FK: C_ID references CUSTOMER (C_ID)

CATEGORY_OF_TICKET_ID references CATEGORY_OF_TICKET(CATEGORY_OF_TICKET_ID)

PAYMENT_METHOD_ID references PAYMENT_METHOD(PAYMENT_METHOD_ID)

FACILITY (FACILITY_ID, FACILITY_DESC, FACILITY_TYPE, FACILITY_CAPACITY, FACILITY_SUBTYPE, LOCATION, TIMESTAMP)

PK: FACILITY_ID
FK: N/A

TICKETS_FACILITY (TICKET_ID*, FACILITY_ID*)

PK: TICKET_ID, FACILITY_ID
FK: TICKET_ID references TICKETS (TICKET_ID)
FACILITY_ID references FACILITY (FACILITY_ID)

RIDES (RIDE_ID, RIDE_NAME, TYPE _OF_RIDE, YEAR_STARTED, DESCRIPTION, MIN_RIDE_HEIGHT(CM), MANUFACTURER, ADDITIONAL FEES)
PK: RIDE_ID
FK: N/A

TICKETS_RIDES (TICKET_ID*, RIDE_ID*)

PK: TICKET_ID, RIDE_ID
FK: TICKET_ID references TICKETS (TICKET_ID)
RIDE_ID references RIDES (RIDE_ID)

**DDL**

```
CREATE TABLE `CUSTOMER` (
  `C_ID` INT,
  `FIRST_NAME` VARCHAR(45),
  `LAST_NAME` VARCHAR(45),
  `EMAIL` VARCHAR(45),
  `GENDER` VARCHAR(10),
  `DOB` DATE,
  `DATE_OF_REG` INT,
  `POSTAL_CODE` VARCHAR(50),
  PRIMARY KEY (`C_ID`)
);

CREATE TABLE `CATEGORY OF TICKET` (
  `CATEGORY_OF_TICKET_ID` INT,
  `CATEGORY_OF_TICKET_DESC` VARCHAR (100),

);
```

```
CREATE TABLE `PAYMENT METHOD` (
  `PAYMENT_METHOD_ID` INT,
  `PAYMENT_METHOD_DESC` VARCHAR (100),
);

CREATE TABLE `TICKETS` (
  `TICKET_ID` INT,
  `PURCHASE_DATE` TIME,
  `PROMO_APPLIED` INT,
  `PURCHASE_MODE` VARCHAR (50),
  `PRICE` FLOAT,
   `C_ID` INT,
  `CATEGORY_OF_TICKET_ID` INT,
  `PAYMENT_METHOD_ID` INT,

   PRIMARY KEY (`TICKET_ID`),
   FOREIGN KEY(`C_ID`) REFERENCES CUSTOMER (C_ID)
   FOREIGN KEY(`CATEGORY_OF_TICKET_ID`) REFERENCES `CATEGORY OF
TICKET` (`CATEGORY_OF_TICKET_ID`)
   FOREIGN KEY(`PAYMENT_METHOD_ID`) REFERENCES `PAYMENT METHOD`
(`PAYMENT_METHOD_ID`)

);

CREATE TABLE `FACILITY` (
  `FACILITY_ID` INT,
  `FACILITY_DESC` VARCHAR(50),
  `FACILITY_TYPE` VARCHAR(50),
  `FACILITY_CAPACITY` INT,
  `FACILITY_SUBTYPE` VARCHAR(45),
  `LOCATION ` VARCHAR(45),
  `TIMESTAMP` TIME,
  PRIMARY KEY (`FACILITY_ID`)
);

CREATE TABLE `PUBLIC TRANSPORATION UTILIZATION` (
  `PUBLICTRANSPORT_ID` INT,
  `TYPE_OF_PUBLIC_TRANSPORT` VARCHAR(100),
  `ACCESSED_THE_PARK` INT,
  `YEAR` INT,
  PRIMARY KEY (`PUBLICTRANSPORT_ID`)
);
```

```
CREATE TABLE `RIDES` (
  `RIDE_ID` INT,
  `RIDE_NAME` VARCHAR(100),
  `TYPE _OF_RIDE` VARCHAR(100),
  `YEAR_STARTED` INT,
  `DESCRIPTION` VARCHAR(100),
  `MIN_RIDE_HEIGHT(CM)` INT,
  `MANUFACTURER` VARCHAR(100),
  `ADDITIONAL FEES` VARCHAR(10),
  PRIMARY KEY (`RIDE_ID`)
);

CREATE TABLE `CUSTOMER_USE` (
  `C_ID` INT,
  `PUBLICTRANSPORT_ID ` INT
   PRIMARY KEY (C_ID, PUBLICTRANSPORT_ID),
   FOREIGN KEY (C_ID) REFERENCES CUSTOMER (C_ID),
   FOREIGN KEY (PUBLICTRANSPORT_ID) REFERENCES PUBLIC TRANSPORATION
UTILIZATION (PUBLICTRANSPORT_ID)
);


CREATE TABLE ` TICKETS_FACILITY ` (
  `TICKET_ID` INT,
  `FACILITY_ID ` INT,
   PRIMARY KEY (TICKET_ID, FACILITY_ID),
   FOREIGN KEY (TICKET_ID) REFERENCES TICKETS (TICKET_ID),
   FOREIGN KEY (FACILITY_ID) REFERENCES FACILITY (FACILITY_ID)
);




CREATE TABLE ` TICKETS_RIDES ` (
  `TICKET_ID` INT,
  `RIDE_ID ` INT,
   PRIMARY KEY (TICKET_ID, RIDE_ID),
   FOREIGN KEY (TICKET_ID) REFERENCES TICKETS (TICKET_ID),
   FOREIGN KEY (RIDE_ID) REFERENCES RIDES (RIDE_ID)
);
```

## ***SECTION-3: Populate the data***

Normalization Process:

- Used excel to perform data cleaning steps on the external data

Populate data

- Using MAMP built-in function to insert normalized data into the dedicated database
- See the attached SQL file to verify the populated of data.

## ***SECTION-4: Queries***

#1

**Business objective**:

To improve the customers' needs, it is important to know which type of transportation utilized to arrive to the park.

what is the trend in mode of transportation from year 2000 to 2006?

```
SELECT public_transport_utilisation.TYPE_OF_PUBLIC_TRANSPORT, CUSTOMER1.GENDER, CUSTOMER1.DOB, CUSTOMER1.C_ID, public_transport_utilisation.YEAR FROM customer_use LEFT JOIN CUSTOMER1 ON customer_use.C_ID=
CUSTOMER1.C_ID LEFT JOIN public_transport_utilisation ON customer_use.PUBLICTRANSPORATION_ID= public_transport_utilisation.PUBLICTRANSPORATION_ID WHERE public_transport_utilisation.YEAR BETWEEN 2000 AND
2006
```

Profiling [ Edit inline ] [ Edit ] [ Explain SQL ] [ Create PHP code ] [ Refresh ]

1 > >> | Show all | Number of rows: 25 Filter rows: Search this table Sort by key: None

+ Options

| TYPE_OF_PUBLIC_TRANSPORT | GENDER | DOB | C_ID | YEAR |
|---|---|---|---|---|
| Bixi | Male | 11/16/2008 | CD0021 | 2000 |
| Metro | Male | 6/21/1990 | CD0022 | 2000 |
| Bus | Female | 1/14/1999 | CD0023 | 2000 |
| Taxi | Female | 7/28/1993 | CD0024 | 2000 |
| Bixi | Male | 5/18/1965 | CD0025 | 2001 |
| Metro | Female | 1/3/1994 | CD0026 | 2001 |
| Bus | Female | 1/18/1974 | CD0027 | 2001 |
| Taxi | Male | 6/14/1977 | CD0028 | 2001 |
| Bixi | Female | 1/4/1969 | CD0029 | 2002 |
| Metro | Male | 7/10/1989 | CD0030 | 2002 |
| Bus | Male | 11/26/2006 | CD0031 | 2002 |
| Taxi | Male | 5/30/2010 | CD0032 | 2002 |
| Bixi | Male | 9/2/2006 | CD0033 | 2003 |
| Metro | Female | 6/20/2000 | CD0034 | 2003 |
| Bus | Male | 2/16/1971 | CD0035 | 2003 |
| Taxi | Male | 9/22/1979 | CD0036 | 2003 |
| Bixi | Female | 3/23/2006 | CD0037 | 2004 |
| Metro | Male | 7/1/1993 | CD0038 | 2004 |
| Bus | Male | 2/2/1967 | CD0039 | 2004 |
| Taxi | Female | 8/19/2008 | CD0040 | 2004 |
| Bixi | Female | 8/28/2002 | CD0041 | 2005 |
| Metro | Female | 5/21/2017 | CD0042 | 2005 |
| Bus | Female | 1/29/1988 | CD0043 | 2005 |
| Taxi | Male | 11/25/1991 | CD0044 | 2005 |
| Bixi | Female | 6/1/2007 | CD0045 | 2006 |

#2

**Business objective**:

If La Ronde Park manager wants to know the maximum and minimum number of people who accessed the park year over year, we can run the query down below. This is key information that facilitates business decision and future expansion of the park.

```
SELECT MAX(ACCESSED_THE_PARK), MIN(ACCESSED_THE_PARK), TYPE_OF_PUBLIC_TRANSPORT, PUBLICTRANSPORATION_ID, YEAR FROM public_transport_utilisation GROUP BY YEAR, PUBLICTRANSPORATION_ID
```

Profiling [ Edit inline ] [ Edit ] [ Explain SQL ] [ Create PHP code ] [ Refresh ]

1 > >> | Show all | Number of rows: 25 Filter rows: Search this table

+ Options

| | MAX(ACCESSED_THE_PARK) | MIN(ACCESSED_THE_PARK) | TYPE_OF_PUBLIC_TRANSPORT | PUBLICTRANSPORATION_ID | YEAR |
|---|---|---|---|---|---|
| Edit Copy Delete | 740000 | 740000 | Bixi | 1 | 1995 |
| Edit Copy Delete | 0 | 0 | Metro | 2 | 1995 |
| Edit Copy Delete | 3009000 | 3009000 | Bus | 3 | 1995 |
| Edit Copy Delete | 0 | 0 | Taxi | 4 | 1995 |
| Edit Copy Delete | 850000 | 850000 | Bixi | 5 | 1996 |
| Edit Copy Delete | 0 | 0 | Metro | 6 | 1996 |
| Edit Copy Delete | 3118000 | 3118000 | Bus | 7 | 1996 |
| Edit Copy Delete | 0 | 0 | Taxi | 8 | 1996 |
| Edit Copy Delete | 0 | 0 | Metro | 10 | 1997 |
| Edit Copy Delete | 3116000 | 3116000 | Bus | 11 | 1997 |
| Edit Copy Delete | 0 | 0 | Taxi | 12 | 1997 |
| Edit Copy Delete | 911000 | 911000 | Bixi | 9 | 1997 |
| Edit Copy Delete | 946000 | 946000 | Bixi | 13 | 1998 |
| Edit Copy Delete | 0 | 0 | Metro | 14 | 1998 |
| Edit Copy Delete | 3121000 | 3121000 | Bus | 15 | 1998 |
| Edit Copy Delete | 0 | 0 | Taxi | 16 | 1998 |
| Edit Copy Delete | 986000 | 986000 | Bixi | 17 | 1999 |
| Edit Copy Delete | 27000 | 27000 | Metro | 18 | 1999 |
| Edit Copy Delete | 3213000 | 3213000 | Bus | 19 | 1999 |

#3

**Business objective**: To target female customers for advertising purpose, we can check gender, public transportation data to isolate females who use metro as a form of transportation. This information can provide us with maximum exposure number and whether to run an ad around the metro station is advisable

```
SELECT CUSTOMER1.GENDER, public_transport_utilisation.TYPE_OF_PUBLIC_TRANSPORT,max(public_transport_utilisation.ACCESSED_THE_PARK) as "Max Exposure" FROM customer_use LEFT JOIN CUSTOMER1 ON customer_use.C_ID
= CUSTOMER1.C_ID LEFT JOIN public_transport_utilisation ON customer_use.PUBLICTRANSPORATION_ID= public_transport_utilisation.PUBLICTRANSPORATION_ID WHERE CUSTOMER1.GENDER = "Female" AND
public_transport_utilisation.YEAR=2008 GROUP BY 1,2
```

Profiling [ Edit inline ] [ Edit ] [ Explain SQL ] [ Create PHP code ] [ Refresh ]

Show all | Number of rows: 25 Filter rows: Search this table

+ Options

| GENDER | TYPE_OF_PUBLIC_TRANSPORT | Max Exposure |
|---|---|---|
| Female | Metro | 88000 |

#4

**Business objective**:

If the La Ronde security manager request how customers arrive to the park since 2006. we can show this by distinct public transport summary with type of transportation and the amount of people who accessed the park. Additionally, the year must be greater than 2005. This information highlights how customers access the park.

```
SELECT DISTINCT (PUBLICTRANSPORATION_ID), TYPE_OF_PUBLIC_TRANSPORT, (ACCESSED_THE_PARK), (YEAR) FROM public_transport_utilisation GROUP BY 1,4 HAVING YEAR >2005
```

Profiling [ Edit inline ] [ Edit ] [ Explain SQL ] [ Create PHP code ] [ Refresh ]

1 > >> | Show all | Number of rows: 25 Filter rows: Search this table Sort by key: None

Options

| | PUBLICTRANSPORATION_ID | TYPE_OF_PUBLIC_TRANSPORT | ACCESSED_THE_PARK | YEAR |
|---|---|---|---|---|
| Edit Copy Delete | 45 | Bixi | 1408000 | 2006 |
| Edit Copy Delete | 46 | Metro | 74000 | 2006 |
| Edit Copy Delete | 47 | Bus | 2833000 | 2006 |
| Edit Copy Delete | 48 | Taxi | 946000 | 2006 |
| Edit Copy Delete | 49 | Bixi | 1527000 | 2007 |
| Edit Copy Delete | 50 | Metro | 79000 | 2007 |
| Edit Copy Delete | 51 | Bus | 2932000 | 2007 |
| Edit Copy Delete | 52 | Taxi | 944000 | 2007 |
| Edit Copy Delete | 53 | Bixi | 1698000 | 2008 |

#5

**Business objective**:

From year 2001 and above if we want to count the mode of transportation and show the total mode of transportation used by the customers.

```
select max(YEAR) as "Year", count(case when TYPE_OF_PUBLIC_TRANSPORT ='Bixi' then 1 end) as bixi_cnt, count(case when TYPE_OF_PUBLIC_TRANSPORT='Metro' then 1 end) as metro_cnt, count(case when TYPE_OF_PUBLIC_TRANSPORT='Bus' then 1 end) as bus_cnt, count(case when TYPE_OF_PUBLIC_TRANSPORT='Taxi' then 1 end) as taxi_cnt, count(*) as total_cnt from public_transport_utilisation where YEAR >2000 group by YEAR
```

Profiling [ Edit inline ] [ Edit ] [ Explain SQL ] [ Create PHP code ] [ Refresh ]

Show all | Number of rows: 25 Filter rows: Search this table Sort by key: None

+ Options

| Year | bixi_cnt | metro_cnt | bus_cnt | taxi_cnt | total_cnt |
|---|---|---|---|---|---|
| 2001 | 1 | 1 | 1 | 1 | 4 |
| 2002 | 1 | 1 | 1 | 1 | 4 |
| 2003 | 1 | 1 | 1 | 1 | 4 |
| 2004 | 1 | 1 | 1 | 1 | 4 |
| 2005 | 1 | 1 | 1 | 1 | 4 |
| 2006 | 1 | 1 | 1 | 1 | 4 |
| 2007 | 1 | 1 | 1 | 1 | 4 |
| 2008 | 1 | 1 | 1 | 1 | 4 |
| 2009 | 1 | 1 | 1 | 1 | 4 |
| 2010 | 1 | 1 | 1 | 1 | 4 |
| 2011 | 1 | 1 | 1 | 1 | 4 |
| 2012 | 1 | 1 | 1 | 1 | 4 |
| 2013 | 1 | 1 | 1 | 1 | 4 |
| 2014 | 1 | 1 | 1 | 1 | 4 |
| 2015 | 1 | 1 | 1 | 1 | 4 |
| 2016 | 1 | 1 | 1 | 1 | 4 |