Assignment 1

This assignment must be completed individually. Submit Word file to EEE drop box. Write your name in the Word file.

1. Given the following four tables:

Customer (CustID, CustName, AnnualRevenue, CustType)
Shipment (ShipmentNumber, CustID, Weight, TruckID, DestinationCity, ShipDate)
Truck (TruckID, DriverName)
City(CityName, Population)

The primary key has a solid underline and foreign key is dashed underlined.
Answer the following queries with SQL script. Use the tables provided above. Please use correct syntax. You do not need MS Access to answer this question. DO NOT submit SQL copied from MS Access.

1) List the **names of drivers** who have delivered shipments for customers with annual revenue over $25 million to cities with populations over 3 million? (1 point)

```
SELECT      DriverName
FROM        Truck, Shipment, City, Customer
WHERE       Truck.TruckID = Shipment.TruckID
        AND City.C
        AND Custo
        AND Custo
        AND City.Population > 3,000,000;
```


2) **How many packages** weighing more than 4 pounds were sent to Los Angeles by customers having annual revenue greater than $500 million?

```
SELECT      COUNT(ShipmentNumber)
FROM        Shipment, Customer
WHERE       Shipment.CustID = Customer.CustID
        AND Shipment.Weight > 4
        AND Shipment.DestirnationCity = "Los Angeles"
        AND Customer.AnnualRevenue > 500,000,000;
```

3) For customers who sent a shipment(s) first to Irvine and later to New York, what is their name and annual revenue? (1 point)

```
SELECT      Customer.CustName, Customer.AnnualRevenue
FROM        Customer, Shipment AS S1, Shipment AS S2
WHERE       Customer.CustID = S1.CustID
        AND S1.CustID = S2.CustID
        AND S1.DestinationCity = "Irvine"
        AND S2.DestinationCity = "New York"
        AND S2.ShipDate > S1.ShipDate;
```

4) List the names of customers who shipped at least 5 packages, each weighing more than 5 pounds to Irvine. (1 point)

```
SELECT      CustName
FROM        Customer, Shipment
WHERE       Customer.CustID = Shipment.CustID
        AND Shipment.DestinationCity= "Irvine"
        AND Shipment.weight > 5
GROUP BY    Customer.CustID
HAVING      COUNT(Shipment.ShipmentNumber) > 4
```

2. **Consider the following table LoanApp.**
   Based on these contingency tables find the following parameters: (2 points)

1) N[Income=low] = 4
2) N[Income=low, Approve=no] = 4
3) N[Liability=high] = 7
4) N[Liability=high, Approve=no] = 6
5) P[Income=low] = 4/14 = 2/7 (0.2857)
6) P[Income=low, Approve=no] = 4/14 = 2/7 (0.2857)
7) P[Approve=no| Income=low] = 1
8) P[Approve=yes| Income=low] = 0
9) P[Liability=high] = 7/14 = 1/2
10) P[Liability=high, Approve=no] = 6/14 = 3/7 (0.4285)
11) P[Approve=no| Liability=high] = 6/7 (0.8571)
12) P[Approve=yes| Liability=high] = 1/7 (0.1428)

3. Continue Question 2:

1) Find the information gain provided by the feature Income on the goal Approve. (2 points)

H(Approve) = H(9/14,5/14) = – 9/14*log2(9/14) – 5/14*log2(5/14) = 0.940
H(Approve|Income=high) = H(2/5,3/5) = – 2/5*log2(2/5) – 3/5*log2(3/5) = 0.971
H(Approve|Income=low) = H(4/4,0/4) = 0
H(Approve|Income=high) = H(3/5,2/5) = – 3/5*log2(3/5) – 2/5*log2(2/5) = 0.971
H(Approve|Income) = 5/14*0.971 + 4/14*0 + 5/14*0.971 = 0.693
I(Approve;Income) = 0.940 – 0.693 = 0.247

H(Approve|Liability=normal) = H(3/7,4/7) = – 3/7*log2(3/7) – 4/7*log2(4/7) = 0.985
H(Approve|Liability=high) = H(6/7,1/7) = – 6/7*log2(6/7) – 1/7*log2(1/7) = 0.592
H(Approve|Liability) = 7/14*0.985 + 7/14*0.592 = 0.789
I(Approve;Liability) = 0.940 – 0.789 = 0.151

2) Find the gain ratio provided by the features Income and Liability on the goal Approve. (2 points)

H(Income) = H(5/14,4/14,5/14)
= – 5/14*log2(5/14) – 4/14*log2(4/14) – 5/14*log2(5/14) = 1.577
H(Liability) = H(7/14,7/14) = – 7/14*log2(7/14) – 7/14*log2(7/14) = 1

G(Approve;Income) = I(Approve;Incom)/H(Income) = 0.247/1.577 = 0.157
G(Approve;Liability) = I(Approve;Liability)/H(Liability) = 0.151/1 = 0.151