ASSIGNMENT 1 - 600.315/415 - Database Systems

Due date: Tuesday, September 29, 2020 at 11:59 PM Baltimore time

Part 1: Database Schema Design (35 points)

Your task is to design a database for a new Baltimore-based variant of Uber, call JHUber. You should support the functions of trip planning, trip pricing, driver (and passenger) evaluation, car amenity assessment and such things as cost/profit assessment for drivers.

For the purposes of this database, you should assume that JHUber's computation model for routes, times and distances is based on pairwise precomputed information between two landmarks and/or two waypoints. A landmark is defined as a major location (e.g. Johns Hopkins Hospital, White Marsh Mall, BWI Airport, etc.) and there are O(500) of these, while waypoints correspond to roughly every street corner (e.g. Pratt and Charles St.), and there are O(10,000) of these. Every landmark is exactly on a waypoint, while not every waypoint corresponds to a landmark. The major distinction is that it's possible to precompute standard times/distances/costs between any two landmarks, while we only have space to store the distances/times between adjacent waypoints in a graph (i.e. the 4 adjacent street corners to the north, south, east and west in most cases).

Between any two landmarks are a series of precomputed routes, corresponding to a list of sequential waypoint pairs that collectively constitute a continuous path between the landmarks.

In addition, for simplification, you should assume that expected traffic loads (and corresponding standard times between waypoints or landmarks) are precomputed and saved for only 3 traffic levels (Low, Medium, High), rather than all different times of a day, and it should be possible from other means for you to compute the expected traffic level on a given date and time (e.g. Saturday 9/26/20 at 5:00 PM) into one of these 3 categories based on the day of the week and whether it's a rush hour time slot or not. You can assume that these traffic relative loads (Low, Medium, High) are the same for all of Baltimore at a given date/time.

From the database, it should be possible to answer the following questions:

(a) What amenities are offered by the JHUber car with license number DXZ211 (e.g. bottled water, iphone charging)?

(b) What is the average age and average total mileage of all cars in the JHUber fleet.

(c) List the names of people who are both drivers and customers and don't drive an electirc vehicle.

(d) List the license number, make, model, production year and owner of the single car in the JHUber database with the highest mileage on the car?

(e) List the license number, make, model, production year and owner of all individual cars in the JHUber database with mileages in excess of 50,000 miles.

(f) List the date, time, nearest waypoint, owner and driver of all JHUber cars which have been in an accident.

(g) List the date, time, nearest waypoint and driver of all drivers who have received a ticket in the database, along with ticket type, infraction details and penalty.

(h) List the make, model, year and mileage of all cars in the database who are driven by someone other than their owners, as well as the driver name and owner name for that car.

(i) For each make+model of vehicle, list the average age, minimum age and maximum age of drivers of that vehicle.

(j) List the average driver review score for each driver in the database who has driven at least one trip, along with his/her name, age and gender.

(k) For each make+model of vehicle, list the average driver review score for all trips involving that vehicle.

(l) List the name, address and gender of the driver with the highest average review score in the database.

(m) List the total miles driven by each driver in the database on 9/24/20.

(n) List the total time driven by each driver in the database on 9/24/20, total miles driven on that day, and average speed on that day (mph = total miles / total hours).

(o) Given the distance driven by each driver on 9/24/20 and the average mpg of their make/model/year of their vehicle and the price of gas on 9/24/20, compute the cost spent by that driver on gas on 9/24/20.

(p) List the total amount spent on gas on 9/24/20, the total driver income on 9/24/20, and total gross profit (income-gasexpenses) for each driver on 9/24/20.

(q) Assuming that JHUber takes a 10% share on all income to JHUber plus a fixed $2.00 booking fee per trip, what is JHUber's corporate income for every day this month.

(r) List the latitude, longitude, zip code and standard name of all landmarks in the database.

(s) List the latitude, longitude, zip code and standard name of all waypoints in the database.

(t) List the expected traffic volume on September 26th, 2020 at 5PM in Baltimore (Low, Medium, High).

(u) List the standard expected distance and standard expected time on a trip between Johns Hopkins Hospital and BWI in LOW traffic volume (which will have been pre-computed).

(v) List the standard distance and expected time between Johns Hopkins Hospital and BWI based on the average of all JHUber actual trip distances/times for that specific trip currently stored in the database.

(w) For customer Russ Taylor with birthdate 2/28/61, what is the expected time, distance and cost of trip tomorrow between landmarks Johns Hopkins Hospital and BWI airport tomorrow (based on stored standard values for that trip).

(x) For customer Russ Taylor with birthdate 2/28/61, what was the difference between the expected time, distance and cost of trip on 9/12/20 between landmarks Johns Hopkins Hospital and BWI airport tomorrow (based on stored standard values for that trip), and the time, distance and cost of his actual trip on that date.

(y) For all trips between landmarks on 9/12/20, list those where the actual distance exceeded the expected standard distance for that trip, as well as the driver for that trip.

(z) List the total distance and total expected time in LOW traffic volume for all trips starting at landmark Johns Hopkins Hospital.

(aa) List the total distance and total expected time in HIGH traffic volume for all trip segments to all waypoints directly adjacent to Johns Hopkins Hospital.

(bb) List the total distance and total expected time in HIGH traffic volume for all possible combined 2 segment trips from the waypoint at Johns Hopkins Hospital, and list the 3 waypoint lat/longs for each of these initial trip segments.

(cc) List all drivers and their car types within 2 miles of Johns Hopkins Hospital.

(dd) List all drivers and their car license numbers that travelled a route that include the Charles and Pratt Street waypoint on 8/24/2020.

(ee) For all drivers in the database, list their most departure landmarks in the database (and the number of total departures from that landmark).

(ff) What is the average driver rating of the most ticketed 10% of drivers.

(gg) List the number of requested rides for each driver, the total number of rides actually driven (after cancellations) and the delta the two.

(hh) List the most requested landmark destination, grouped by day of the week traveled.

(ii) List the top 10% of vehicle types reporting lost hours due to unscheduled maintenance, including the average number of hours lost per month.

(jj) List the driver with the highest total gross income in the database (passenger fare minus JHUber share).

(kk) List the total amount spent on gas and maintenance for each car in the database.

(ll) List the most profitable driver in the database (total gross income minus total spent on gass and maintenance for the cars that they drive).

(mm) List the expected time for all trips between landmarks that leave from Johns Hopkins Homewood Campus in HIGH traffic volume, including time, distance and destination landmark.

(nn) For all trips between any 2 landmarks, find the difference between the costs of the trip for LOW and HIGH traffic volumes.

(oo) List the drivers who drove more than 50,000 miles (combined total from all cars driven) but have not received a ticket.

(pp) List the customers who have been driven by the driver(s) with the highest review score and the driver(s) with the lowest review score.

(qq) List the drivers who drive only during the weekends or after 6PM.

(rr) Find the average driver review score for cars with each type of amenity (for all trips using that car)

(ss) List the drivers who have visited (either arrived or departed from) every landmark in the database.

(tt) What is the average time (between low, medium and high traffic times) to get from Johns Hopkins Homewood campus to Johns Hopkins Hospital?

(uu) Of the cars with a production year after 2012, what model of car that makes the most money during low traffic levels and has a driver who lives in Baltimore City?

(vv) What is the average difference in pricing between low and high traffic levels for a ride from BWI Airport to Homewood campus in a Chevy Volt (any year/owner)?

(ww) What is the license number and make of the most expensive ride you can take during high traffic levels from Camden Yards to Eddies Grocery Store?

(xx) Find the average number of miles per trip driven by all drivers working on 9/24/20 and the gas mileage of the car they drive.

(yy) What is the most popular JHUber car by gender of driver?

(zz) What is the longest path between any two landmarks in the database?

(aaa) What is the most often visited landmark in the database?

(bbb) Which customer has travelled the most miles?

(ccc) Which make+model+year of car in the JHUber fleet has the best gas mileage?

(ddd) Find the average difference between actual trip time and expected trip time for all 3 traffic conditions(High, Medium and Low).

Simplifying assumptions:

You can assume that a driver can drive multiple cars, a person can own multiple cars, an individual car can be driven by multiple drivers, but a car can only be owned by one person.

1.1 (*20 points*) Design the database using the entity-relationship database model and draw it. Your design should minimize repetitions of information, and distinguish between generic and specific/individual instantiations of concepts as discussed in the design exercise in class. Be sure to underline the primary keys, and also specify all mapping constraints and participation constraints using the (lower_bound,upper_bound) syntax such as *(0,1)* or *(1,N)*, as well as obligatory participation via a double line. You may *also* represent participation constraints via arrows (e.g. $\leftrightarrow$, $\rightarrow$, $\leftarrow$, $\rightarrow$ or $\Rightarrow$), redunantly to the more precise numeric bounds above, but this is not necessary.

You should *very* briefly justify any unusual or potentially controversial design decisions you make. Do *not* spend much time on such notes.

This section will be graded on aesthetics and completeness as well as correctness.

1.2 (*10 points*) Represent this database design using the relational model. You should use a tabular notation and include at least one row of sample values for each relation.

You should *very* briefly justify any unusual or potentially controversial decisions you make in the conversion process. Do *not* spend much time on such notes.

1.3 (*5 points*) Write relational algebra expressions for queries *c, f, h, y, ss, ccc* based on your relational database design. 601.315 students should answer questions for only 3 queries of their choice.

Part 2: Relational Algebra and Relational Calculus (65 pts.)

Consider the following hypothetical database schema. Suppose all bars in the US have a unique bar license number (BNO) and each drinker is identified by a unique drivers' license number (DLicNo). Given that this is the year 2020, all bars implement contact tracing, with a record of all bar visits and COVID status for all visitors to a bar. For example, every time a drinker represented by DLicNo goes to a bar represented by BNO, the information is recorded in the database. The number of times a drinker visits a particular bar can be obtained by examining the VISIT relation. Likewise, given that this is 2020, given a variety of monitoring techniques (e.g. face-recognition-based and phone/gps/social-media monitoring), the database also stores information about a bar visitors likes, purchases, and activities. The relation LIKES represents all the beers that a particular drinker likes and the relation SERVES represents all the beers a particular bar serves.

BAR	BNO	BarName	BCity	BState
	L22174	*Murphy's*	*Towson*	*MD*
	L31927	*Joe's*	*Lutherville*	*MD*
	L59871	*BatBar*	*Georgetown*	*DC*

DRINKER	DLicNo	DName	DCity	Age	Phone	PoliticalParty
	AK117229	*Donald Trump*	*New York*	*75*	*201-555-6666*	Republican
	UU761326	*Melania Trump*	*Slovenia*	*51*	*201-555-9999*	Independent
	ZM193312	*Joe Biden*	*Wilmington*	*76*	*215-555-7777*	Democratic
	MD891129	*Mike Pence*	*Indianapolis*	*68*	*201-555-4321*	Republican
	YU134618	*Ivanka Trump*	*New York*	*41*	*201-555-0001*	Libertarian

VISIT	DLicNo	BNO	DateOfVisit
	AK117229	*L22174*	*2020-09-12*
	MD891129	*L59871*	*2020-11-05*
	AK117229	*L59871*	*2020-10-03*

LIKES	DLicNo	BeerName
	AK117229	*Bud Lite*
	AK117229	*Rolling Rock*
	MD891129	*Sam Adams*

SERVES	BNO	BeerName
	L22174	*Bud Lite*
	L59871	*Bud Lite*
	L59871	*Rolling Rock*

BEER_PURCHASE	DLicNo	BNO	BeerName	DateOfPurchase
	MD891129	*L59871*	*Coors Lite*	*2020-11-05*
	YU134618	*L22174*	*Bud Lite*	*2020-11-05*

COVID_DIAGNOSIS	DLicNo	EstimatedStartDate	DateOfDiagnosis	EstimatedEndDate
	YU134618	*2020-05-06*	*2020-05-08*	*2020-05-20*
	AK117229	*2020-04-01*	*2020-04-05*	*2020-04-15*
	YU134618	*2020-08-06*	*2020-08-09*	*2020-08-20*

COVID_VACCINE	DLicNo	DateOfVaccine	Manufacturer
	UU761326	*2020-10-15*	*Oxford*
	ZM193312	*2020-10-03*	*Sputnik5*
	YU134618	*2020-10-02*	*Merk*

Students in 601.315 should answer queries 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 2.14, 2.15, 2.16, 2.17, 2.18, 2.19, 2.23 in the *relational algebra*.

Students in 601.315 should *also* answer *any* 4 queries of your choice from 2.1, 2.2, 2.3, 2.4, 2.9, and 2.10 in the *relational calculus*.

Students in 601.415 and 601.615 should answer **all** 23 of the queries below in the *relational algebra*.

Students in 601.415 and 601.615 should *also* answer queries 2.1, 2.2, 2.3, 2.4, 2.9, and 2.10 in the *relational calculus*.

2.1 List the name and political party for all individuals who have visited a bar on the same day that Ivanka Trump has visited that bar.

2.2 List the names of bars in Maryland that are *not* in Baltimore **and** do not serve Bud Lite.

2.3 List the names of all people under 30 who have visited at least one bar in Georgetown and like Bud Lite *and* do not like Miller Lite.

2.4 List the name and age of everyone who has visited at least one bar that Donald Trump has visited.

2.5 List the names and ages of all people who have visited every bar in Towson.

2.6 List the name and birthdate for all individuals who have neither visited a bar nor suffered from COVID.

2.7 List the names and ages of people who have visited every bar that Donald Trump has visited and have never visited a bar that Joe Biden has visited.

2.8 List the names of people who have never drunk a beer named for them (e.g. "Sam Adams" drinking a beer called "Sam Adams"), but have visited at least 1 bar named for them.

2.9 List the name of every bar that serves a beer that Donald Trump doesn't like.

2.10 List the name of every bar in Towson that serves no beer that is served in a Bar in Timonium.

2.11 List the name of all beers that both Donald Trump and Ivanka Trump like and are served at the same bar in the database (i.e. a bar where both people could order a beer that they like).

2.12 List the name, city and state of the bar that serves the greatest number of different beers.

2.13 List the name and age of the drinker that likes the fewest number of different beers but likes at least one beer.

2.14 List the name and Political Party of all drinkers who like no beer that Donald Trump or Mike Pence likes.

2.15 List the name and Political Party of all drinkers who like every beer that Donald Trump likes.

2.16 List the names of all beers purchased by people while they were suffering from COVID.

2.17 List the name of every bar that Donald Trump has visited more than once.

2.18 List the name and phone number of each person who visited a bar on the same date that Ivanka Trump visited the bar during the date window when she likely had covid.

2.19 List the name and phone number of each person who visited a bar on the same date as **any** COVID-19 sufferer who visited the bar during the date window when they likely had covid.

2.20 For all people who have visited a bar during the time window they have been suffering from COVID-19, list a 4-tuple of the name of that individual, the name of the bar that they visited, the estimated start date of their COVID and estimated end date of their COVID. If an individual has visited 3 bars while suffering from COVID, then you should list 3 separate tuples for that person (one for each bar).

2.21 List the names and ages of people who have visited at least every bar that Joe Biden has visited, and has visited all of these bars the identical number of times that Joe Biden has visited the bar.

2.22 List the name and age and political party of all drinkers who were vaccinated for COVID and then subsequently tested positive for COVID on a different date.

2.23 List the name and age of drinkers who have purchased all of the beers that are served by BatBar and also purchased them at BatBar