Part I - Querying the database to answer a key business question

BuildZoom representatives record notes for every conversation with a contractor. When representatives check in with contractors, i.e. when they initiate the call, they indicate that it was a check-in call by including the term “#checkin” in the notes.

Questions:

1. Using the “database” below, write an **automated** script or query that reports what share of contractors received at least one check-in call in 2016.
2. Using the “database” below, write an **automated** script or query that generates a list of contractors, by name, who received at least one check-in call in 2016.

If you are familiar with it, we prefer that you write the query in MySQL.

contractors	
contractor_id	business_name
88766373	ABC Construction
24963459	DEF Construction
78053782	GHI Construction
87753153	JKL Construction
90096012	MNO Construction
90096013	PQR Construction
23909579	STV Construction
90096014	WXY Construction
90096015	Z Construction

notes		
contractor_id	note	date
88766373	blah blah blah	10/18/2016
24963459	#checkin	4/24/2016
78053782	blah blah blah	4/24/2016
87753153	blah blah blah	11/21/2013
88766373	blah blah blah blah blah blah #checki	4/22/2016
24963459	blah blah blah	4/23/2016
78053782	blah blah blah	4/24/2016
87753153	blah blah blah	4/25/2016
88766373	blah blah blah #checkin	10/31/2015
24963459	blah blah blah	11/1/2015
78053782	blah blah blah	11/2/2015
87753153	blah blah blah	1/29/2016
88766373	blah blah blah	10/30/2015
24963459	blah blah blah #checkin	4/24/2016
88766373	blah blah blah	11/1/2015
24963459	blah blah blah	11/2/2015
78053782	blah blah blah blah blah blah checking to see if they will need projects	11/3/2015
87753153	blah blah blah	11/4/2015
88766373	blah blah blah #checkin	11/5/2016
24963459	blah blah blah	11/6/2015

Part II - Using analytics to optimize outreach to contractors

BuildZoom has signed up 100,000 contractors, more than enough to support our multi-billion dollar portfolio of projects. Not all of these contractors are actively engaged with BuildZoom. Therefore, we try to engage them by phone, email, text message or through the contractor dashboard on our website.

We currently evaluate contractors' engagement using three metrics:

- Opt-in Rate (OIR): the share of projects projects offered to contractors in which they express interest.
- Bid Rate (BR): the share of projects that contractors opted in to, in which they submit a written bid (i.e. proposal) on our website.
- Win Rate (WR): the share of projects that contractors opted in to, in which they are hired.

Let $j \in J$ denote projects and $i \in I$ denote contractors.

$L_{ij} = 1$ if contractor i was offered project j; $L_{ij} = 0$ otherwise.

$O_{ij} = 1$ if contractor i opted in to project j; $O_{ij} = 0$ otherwise.

$B_{ij} = 1$ if contractor i bid on project j; $B_{ij} = 0$ otherwise.

$H_{ij} = 1$ if contractor i was hired for project j; $H_{ij} = 0$ otherwise.

Questions:

1. Consider the following ways of expressing the opt-in rate:

$$OIR^1 = |\{j \in J | \sum_i O_{ij} > 0\}| / |J|$$

$$OIR^2 = |I|^{-1} \sum_{i \in I} \frac{\sum_{j \in J} O_{ij}}{\sum_{j \in J} L_{ij}}$$

$$OIR^3 = \frac{\sum_{i \in I} \sum_{j \in J} O_{ij}}{\sum_{i \in I} \sum_{j \in J} L_{ij}}$$

   What does each of these expressions capture? In what context(s) would it be more appropriate to use one of these measures rather than another, and why?

2. How would you evaluate the effectiveness of one or more engagement tactics, e.g. text and email messaging? You may speculate as to the nature of the data available, but make your assumptions explicit.
   a. Present the method(s) of evaluation as explicitly as possible, providing both a mathematical and a verbal representation.
   b. What output would the method(s) deliver, and how would you interpret it?
   c. What evaluation results do you anticipate, intuitively?

    d. What steps would be involved in implementing the evaluation, e.g. coding an algorithm, running an experiment, conveying the findings to others? How long would you expect each step to take you?
    e. If you were asked to halve the total evaluation time, where would you compromise?

3. What alternative engagement metric(s) would you recommend using?
    a. Define it precisely and explain what it captures.
    b. Suggest a way of measuring it, e.g. what kind of data would be necessary and how would it be obtained.
    c. Explain why the metric is useful and how its addition would improve the set of engagement metrics.

Part III - Qualifying incoming service requests

BuildZoom would benefit from being able to predict which of the homeowners submitting service requests on our website are more likely to ultimately hire a contractor through us. The reasons that a homeowner may ultimately fail to hire through us are varied. Often, such failure reflects a failure to hire at all, through BuildZoom or otherwise.

This file contains a dataset of characteristics for ~6.6K homes associated with service requests submitted to BuildZoom. The first column, service_request_won, indicates whether a contractor was ultimately hired for the job through BuildZoom. The remaining columns are drawn from an external dataset, whose definitions may be found here.

Questions:

1. Using the information given, which home features associated with service requests are the most useful in predicting a hire through BuildZoom?
2. How would you address the task of predicting the probability of hire through BuildZoom using these data if you were given:
    a. One day to complete the task.
    b. Two weeks to complete the task.

    Be as precise as possible.