Overview of the data

funnel.csv
Data about events occurring on Affirm's checkout product (try it out at a merchant like Casper to get a sense for the flow)

- merchant_id: Unique identifier for the merchant (links to merchants.csv)
- user_id: Unique identifier for the user (only populated after the user logins when 'Loan Terms Run' action takes place)
- checkout_id: Unique identifier for a given checkout (links to loans.csv)
- action: Name of the event, can be "Checkout Loaded" = checkout page was loaded, "Loan Terms Run" = user applied for a loan, "Loan Terms Approved" = user was approved for a loan, "Checkout Completed" = user took the loan for which they were approved
- action_date: Date when the event happened

loans.csv
Data on each loan from the 'Completed Checkout' action

- merchant_id: Unique identifier for the merchant
- user_id: Unique identifier for the user
- checkout_id: Unique identifier for a given checkout
- checkout_date: Date when checkout was completed
- loan_amount: total amount of the loan
- user_first_capture: first date the user took out a loan with Affirm (only populated if repeat Affirm user)
- user_dob_year: year the user was born
- loan_length_months: length of the loan in months
- mdr: merchant discount rate (transaction rate charged to the merchant for each loan)
- apr: annual percentage rate (interest rate charged to the user)
- fico_score: score that measures a user's risk, higher score means less risk (ranges from 300-850)
- loan_return_percentage: The return Affirm saw on the loan (negative values mean the loan was not paid back in full)

merchants.csv
Data on each merchant that integrates Affirm's checkout product

- merchant_id: Unique identifier for the merchant
- merchant_name: Name of the merchant
- category: The merchant's industry

Questions

1. **Please review the integrity of the data. Do you notice any data anomalies? If so, please describe them.**

2. Calculate conversion through the funnel by day such that the data structure is:

Date	num_loaded	num_applied	num_approved	num_confirmed	application_rate	approval_rate	confirmation_rate
2016-05-01	100	80	60	30	0.8	0.75	0.50
2016-05-02	120	90	81	63	0.75	0.90	0.78

2a. Please provide a SQL query you used or would use to calculate the application rate by merchant category (the merchant's industry provided in *merchants.csv*).

3. Provide a set of recommendations on how to improve our business or product based on the attached dataset (assume we have roughly the same market penetration in each so that saturation isn't a concern and assume revenue to Affirm = (mdr + loan_return_percentage). Please put together a 3-page PowerPoint presentation to the executive team with your recommendation.

This is intended to be fairly open-ended - there's no right or wrong answer. We're more concerned with your approach and the insights you uncover.

4. Choose one of the recommendations/insights you uncovered (in #3) and outline one experiment you would like to run to test your suggested product/business recommendation. Please state your hypothesis, describe how you would structure your experiment, list your success metrics and describe the implementation.

5. Let's assume that the experiment you ran (in #4) proved your hypothesis was true. How would you suggest implementing the change on a larger scale? What are some operational challenges you might encounter and how would you mitigate their risk?