MoneyLion Data Scientist Assessment Data Dictionary:

In the data.zip file you will find 3 files:

1. loan.csv
2. payment.csv
3. clarity_underwriting_variables.csv

1. loan.csv:

In this file there are 18 columns:

1. loanId:
    a. This is a unique loan identifier. Use this for joins with the payment.csv file
2. anon_ssn:
    a. This is a hash based on a client's ssn. You can use this as if it is an ssn to compare if a loan belongs to a previous customer.
3. payFrequency:
    a. This column represents repayment frequency of the loan:
        i. B is biweekly payments
        ii. I is irregular
        iii. M is monthly
        iv. S is semi monthly
        v. W is weekly
4. apr:
    a. This is the loan apr %
5. applicationDate:
    a. Date of application (start date)
6. originated:
    a. Whether or not a loan has been originated (first step of underwriting before loan is funded)
7. originatedDate:
    a. Date of origination- day the loan was originated
8. nPaidOff:
    a. How many MoneyLion loans this client has paid off in the past
9. approved:
    a. Whether or not a loan has been approved (final step of underwriting before a loan deposit is attempted)
10. isFunded:
    a. Whether or not a loan is ultimately funded. –a loan can be voided by a customer shortly after it is approved, so not all approved loans are ultimately funded.
11. loanStatus:
    a. Current loan status. Most are self explanatory. Below are the statuses which need clarification:
        i. Returned Item: missed 1 payment (but not more), due to insufficient funds
        ii. Rejected: Rejected by automated underwriting rules – not by human underwriters
        iii. Withdrawn Application – application abandoned for more than 2 weeks, or is withdrawn by a human underwriter or customer

        iv. Statuses with the word “void” in them mean a loan that is approved but cancelled. (One reason is the loan failed to be debited into the customer’s account).
12. loanAmount:
    a. Principal of loan – for non-funded loans this will be the principal in the loan application
13. originallyScheduledPaymentAmount:
    a. This is the originally scheduled repayment amount (if a customer pays off all his scheduled payments, this is the amount we should receive)
14. state
    a. Client’s state
15. Lead type
    a. The lead type determines the underwriting rules for a lead.
        i. bvMandatory: leads that are bought from the ping tree – required to perform bank verification before loan approval
        ii. lead: very similar to bvMandatory, except bank verification is optional for loan approval
        iii. california: similar to (ii), but optimized for California lending rules
        iv. organic: customers that came through the MoneyLion website
        v. rc_returning: customers who have at least 1 paid off loan in another loan portfolio. (The first paid off loan is not in this data set).
        vi. prescreen: preselected customers who have been offered a loan through direct mail campaigns
        vii. express: promotional “express” loans
        viii. repeat: promotional loans offered through sms
        ix. instant-offer: promotional “instant-offer” loans
16. Lead cost
    a. Cost of the lead
17. fpStatus
    a. Result of the first payment of the loan:
        i. Checked – payment is successful
        ii. Rejected – payment is unsuccessful
        iii. Cancelled – payment is cancelled
        iv. No Payments/No Schedule – loan is not funded
        v. Pending – ACH attempt has been submitted to clearing house but no response yet
        vi. Skipped – payment has been skipped
        vii. None – No ACH attempt has been made yet – usually because the payment is scheduled for the future

18. clarityFraudId:
    a. unique underwriting id. Can be used to join with columns in the clarity_underwriting_variables.csv file

Every row represents an accepted loan application/ successfully funded loan.

Missing values can exist. Some fields are only implemented after the loan application was made.

2. payment.csv:

9 columns in this file:

1. loanId:
    a. This is a unique loan identifier. Use this for joins with the loan.csv file
2. isCollection
    a. A loan can have a custom made collection plan if the customer has trouble making repayments as per the original schedule. TRUE means the payment is from a custom made collection plan.
3. installmentIndex
    a. This counts the nth payment for the loan. First payment is 1, 2nd payment is 2 and so on.
    b. This index resets for collection payment plans. So some loans can have 2 payments with the same installmentIndex. One from the regular plan and one from the collection plan.
4. paymentdate
    a. Effective of payment
5. prinicpal
    a. principal component of the payment
6. fees
    a. Fee/ interest amount of the payment
7. paymentAmount
    a. Total amount of the payment
    b. Usually equals to fees + principal
8. paymentStatus
    a. Checked – payment is successful
    b. Rejected – payment is unsuccessful
    c. Cancelled – payment is cancelled
    d. Pending – ACH attempt has been submitted to clearing house but no response yet
    e. Skipped – payment has been skipped
    f. None – No ACH attempt has been made yet – usually because the payment is scheduled for the future
    g. Rejected awaiting retry – retrying a failed ACH attempt.
9. paymentReturnCode: these are ACH error codes to explain why the payment failed. You can find more information about this at the end of this document, or visit the following link:
    a. https://www.vericheck.com/ach-return-codes/

Each row in this file represents an ACH attempt (either scheduled for the future or has elapsed in the past) associated to the loan.

3. clarity_underwriting_variables.csv:

This file contains variables returned from a data provider that is used to underwrite loans.

Each row represents an underwriting report. The underwritingid in the last column can be used to join with the clarityFraudId column in loan.csv.

Not all rows in loan.csv has a corresponding row in clarity_underwriting_variables.csv. The availability of the clarity variables depends on the underwriting flow for the lead.

The meanings of the columns are described in the files in dictionaries.zip. The csv and xlsx files are equivalent, just different file formats.)

Appendix:

ACH return codes:

R01	Insufficient Funds
R02	Account Closed
R03	No Account/Unable to Locate Account
R04	Invalid Account Number
R05	Unauthorized Debit Entry
R06	Returned per ODFI's Request
R07	Authorization Revoked by Customer (adjustment entries)
R08	Payment Stopped or Stop Payment on Item
R09	Uncollected Funds
R10	Customer Advises Not Authorized; Item Is Ineligible, Notice Not Provided, Signatures Not Genuine, or Item Altered (adjustment entries)
R11	Check Truncation Entry Return
R12	Branch Sold to Another DFI
R13	RDFI not qualified to participate
R14	Representative Payee Deceased or Unable to Continue in that Capacity
R15	Beneficiary or Account Holder (Other Than a Representative Payee) Deceased
R16	Account Frozen
R17	File Record Edit Criteria (Specify)
R20	Non-Transaction Account
R21	Invalid Company Identification
R22	Invalid Individual ID Number
R23	Credit Entry Refused by Receiver
R24	Duplicate Entry
R29	Corporate Customer Advises Not Authorized
R31	Permissible Return Entry (CCD and CTX only)
R33	Return of XCK Entry