# РОЗДІЛ 3 ПОБУДОВА СКОРИНГОВОЇ МОДЕЛЕЙ

3.1 Характеристика вхідних даних

Щоб побудувати скорингову модель були відібрані реальні дані з бази даних «Альфа-Банку», який за розмірами активів та часткою на роздрібному ринку входить в десятку найбільших банків України.

Загальні активи Альфа-Банку, на кінець 2015 року, становили понад 42 мільярди гривень. Роздрібна мережа обслуговування включає в себе близько 100 відділень та 200 банкоматів. За початкові дані були взяті репрезентативна вибірка банку АТ «Альфа-Банк». Вибірка складається із 20 змінних:

1. Стан поточного рахунку
2. Тривалість кредиту
3. Кредитна історія
4. Мета кредиту
5. Сума кредиту
6. Ощадний рахунок / облігації
7. Стаж поточної роботи
8. Ставка внеску (у відсотках від наявного доходу)
9. Сімейний статус і стать
10. Відповідальна особа
11. Нинішня резиденція
12. Тип власності
13. Вік
14. Заявник має інший кредит на розстрочку
15. Тип власності на дім
16. Кількість кредитів у цьому банку
17. Тип зайнятості
18. Кількість людей, які несуть відповідальність за кредит
19. Номер телефону(Є/ Немає)
20. Іноземний працівник (Так / Ні)

Перевіримо наявність відсутніх значень:

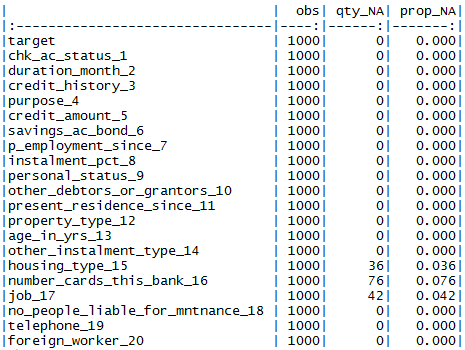


Рис.3.1.1. Розподіл пропущених значень у вибірці

Колонка «obs» - всого кількість спостережень у змінній;

«qty\_NA» - к-сть пропущених значень;

«prop\_NA» - відсоткове відношення кількості пропущених значень до загальної кількості спостережень у вибірці.

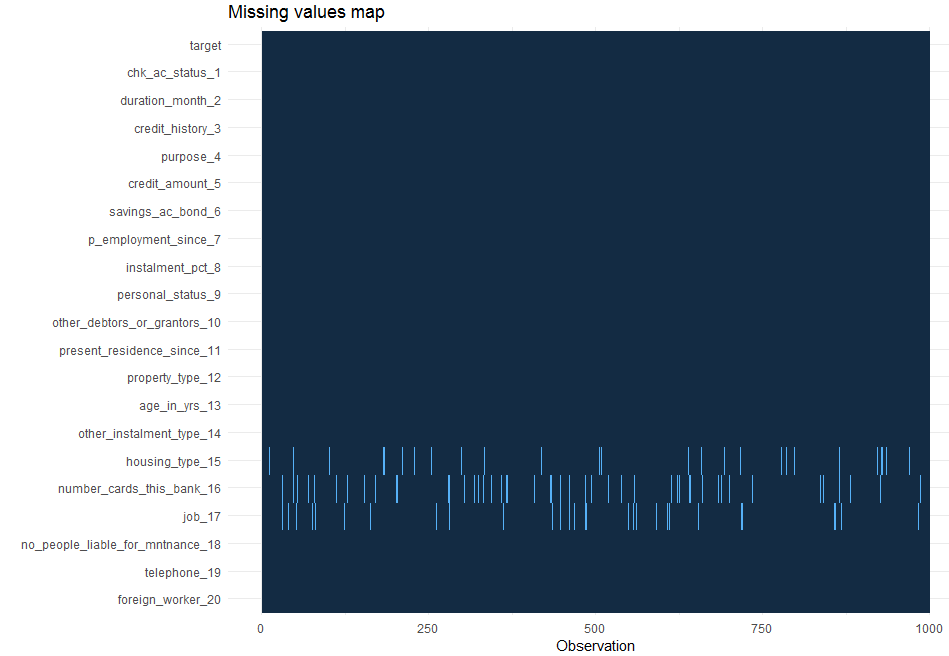
Охарактеризуємо розподіл пропущених значень:

Рис.3.1.2. Мапа пропущених значень

Пропуски у даних розподілені рандомно, тобто тип пропущених значень – MCAR (Missing Completely At Random). Це механізм формування пропусків, при якому ймовірність пропуску для кожного запису набору однакова.

Переконаємося у правильності припущень.

Охарактеризуємо тип даних змінних:

