Part -a

URL of the webpage: https://archive.ics.uci.edu/ml/machine-learning-databases/00529/

URL for the site:
https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset.

Part -b

Brief description of data set:

--The data set is used for the early-stage diabetes risk prediction, it is one CSV file containing all the information

-- Objects: we have a data set of size and length (520, 17)

--Attributes: We have 17 attributes for the above data set. which is a combination of different types namely Gender(objects), age(Integer), others are binary either yes or no, and in the last column name class it is positive/negative (objects)

Please find the below information related to the given dataset, which defines how many objects and attributes are present in the dataset and their types,

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 520 entries, 0 to 519
Data columns (total 17 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   Age                 520 non-null    int64 
 1   Gender              520 non-null    object
 2   Polyuria            520 non-null    object
 3   Polydipsia          520 non-null    object
 4   sudden weight loss  520 non-null    object
 5   weakness            520 non-null    object
 6   Polyphagia          520 non-null    object
 7   Genital thrush      520 non-null    object
 8   visual blurring     520 non-null    object
 9   Itching             520 non-null    object
 10  Irritability        520 non-null    object
 11  delayed healing     520 non-null    object
 12  partial paresis     520 non-null    object
 13  muscle stiffness    520 non-null    object
 14  Alopecia            520 non-null    object
 15  Obesity             520 non-null    object
 16  class               520 non-null    object
dtypes: int64(1), object(16)
```

Part -c:

- Early diabetes identification is usually preferred for a clinically significant result due to the existence of a relatively extended asymptomatic phase.
- From the data set we have based on the symptoms like sudden weight loss, Irritability, obesity, etc a person can be classified whether he has diabetes or not
- Only by careful evaluation of both common and uncommon symptoms, which can be identified at various stages from disease onset to diagnosis, is an early diagnosis of diabetes possible.

Part -D:-

- This knowledge is quite useful, if in the future if a doctor gets a patient and he has some of the symptoms lets say 6 symptoms mentioned in the table, then he has a high risk of getting diabetes or he is already suffering from that
- The applications of data-mining techniques in the selected articles were useful for extracting valuable knowledge and generating new hypothesis for further scientific research/experimentation and improving health care for diabetes patients.
- The results could be used for both scientific research and real-life practice to improve the quality of health care for diabetes patients.