## Data Analyst

## Interview ETL problem

*Objective*

Write an original ETL code that combines files and identifies cases that have a low blood pressure based on logic below.

A successful project includes:

1. Original code file (Python)
2. Final report (excel or csv)

*Steps*

1. Identify cases during which blood pressure dropped below the norm for the age (see below) for 14 continuous minutes or longer.

Assume, that the PERSON\_ID is the identifier for the patient, and SERVICE\_DATE is the date of the surgery that they had. Surgeries don’t span over 1 day. The Age is given for that patient, for the surgery date. The Blood pressure is only taken during the surgery duration.

*If the child reached 44 months, systolic blood pressure is considered low at 55 mmHg and below. Before 44 months of age, 46 mmHg and below is considered low.*

*Example:*

*Demographic data:*

|  |  |  |
| --- | --- | --- |
| *PERSON\_ID* | *SERVICE\_DATE* | *AGE\_MONTHS* |
| *1* | *1/12/2016* | *40* |

*Blood Pressure data:*

|  |  |  |  |
| --- | --- | --- | --- |
| *Row Number* | *PERSON\_ID* | *TIME* | *SYSTOLIC\_BLOOD\_PRESSURE* |
| *1* | *1* | *1/12/2016 07:05* | *32* |
| *2* | *1* | *1/12/2013 07:06* | *54* |
| *3* | *…* |  |  |

*Since the patient is 40 months old at the time of the surgery, their threshold is 46 mmHg, hence, only row 1 would be considered low for 1 minute.*

1. The final report should contain the Person ID, Service date, and duration (in minutes) of the period with low blood pressure.

*Source Files*

Demographics.csv

BloodPressure.csv

*Additional considerations:*

The blood pressure thresholds are not clinically correct, and are created for this problem only.