Upon to two distinctive constraint definitions in my advance project 2 report, checking those hypothetical terms in real life data is decided. To be able to observe interesting patterns, a big data set with 2-3 years production orders is agreed to be investigated through time windows.

I and Daniel started to discuss about the relevant features to be considered in this data set. At final stage below given SQL query was generated to pull the data set from the SMS database. The resultant data set consists of 459203 rows and 15 columns. First two columns and 4^{th} column features: ROS.R_OS_ID, ROS.PRODUCTION_LINE_NAME, and ROS.REFERENCE_DATE come from "Reporting data: Operation step" table. 3^{rd} column feature SEQUENCE_ID is actual casting sequence ID from the table "Reporting data: additional data of CCM (explain this)". 5^{th}., 6^{th}., 7^{th}. and 14^{th}. SLAB.PIECE_ID, SLAB.MATERIAL_ID, SLAB.MOLD_WIDTH, and SLAB.EXIT_TEMP come from "Reporting data: additional data of CCM which are slab related". Rest of the columns: MAT.WIDTH, MAT.THICKNESS, MAT.WEIGHT, MAT.LENGTH, MAT.HEAR_ID, MAT.STEEL_GRADE_ID_INT, and MAT.SLAB_TRANSITION come from "Material ; For slabs, coils, plates and heats" table.

```
SELECT  ros.r_os_id, ros.production_line_name, ccm.sequence_id,
        ros.reference_date, NVL(TO_CHAR(slab.piece_id),'NA') piece_id,
        NVL(TO_CHAR(slab.material_id),'NA') material_id,
        NVL(TO_CHAR(slab.mold_width),'NA') mold_width,
        NVL(TO_CHAR(mat.width),'NA') width,
        NVL(TO_CHAR(mat.thickness),'NA') thickness,
        NVL(TO_CHAR(mat.weight),'NA') weight, NVL(TO_CHAR(mat.length),'NA')
        length, NVL(TO_CHAR(mat.heat_id),'NA') heat_id,
        NVL(TO_CHAR(mat.steel_grade_id_int),'NA') steel_grade_id_int,
        NVL(TO_CHAR(slab.exit_temp),'NA') exit_temp,
        NVL(TO_CHAR(mat.slab_transition),'NA') slab_transition

FROM    L3MAIN.r_os ros
        LEFT JOIN L3MAIN.r_ccm ccm ON ros.r_os_id = ccm.r_os_id
        LEFT JOIN L3MAIN.r_ccm_slab slab ON ros.r_os_id = slab.r_os_id
        LEFT JOIN L3MAIN.r_mat mat ON ros.r_os_id = mat.r_os_id

WHERE   sequence_id IS NOT NULL;
```

Converting strings to numbers and correction for punctuation marks between digits were performed, null values (NA) were converted into 0 values in the beginnning of data cleaning process. After completing minor stages, some preconditions were generated as below to be able to manipulate data columns. Steel density is considered between 7.00 x $10^(-6) kg/mm^(3) and 8.50x10^(-6) kg/mm^(3). Width varies between 800-2000 mm. Thickness varies between 40-90 mm. Weight varies between 2669-26690 kg. Length unit is mm.$

Starting to modify width, thickness, and weight values corresponding to thickness values with 2 digits.

The data set has below given shape just before starting to analysis. Weight

Zero Rows: 10484 Thickness + Width Zero Rows: 61320 Usable Rows: 397883
Time Windows Generation by Data Partitioning