In the attached zip file, please find a .ipynb file: job-a-thon-workbook.ipynb.

The notebook has 3 main sections:

1. File load functions

2. Data utilities

3. Model Checking functions

Ten models have been developed in this notebook ( each represented in a subsection in the notebook ). Three models – m3, m4 and m6 are invalid because they make use of moving averages and would fail ( because there is no demand value in the test file ).

All models make use of linear regession and were developed by splitting the training set in the raio of 4:1 ( 4 for training, 1 for validating ).

The final model for submission is m10 which was derived from the following features:

|  |  |  |
| --- | --- | --- |
| Ser. No. | Name | Description |
| 1. | hour | Hour |
| 2. | year | Year |
| 3. | month | Month |
| 4. | day | Day |
| 5. | day\_of\_week | Day of week ( 0=Monday, 7=Sunday ) |
| 6. | hour\_type | Blocks for day ( hblk1, hblk2, hblk3, hblk4) |
| 7. | month\_type | Blocks for month ( mblk1, mblk2, mblk3, mblk4 ) |
| 8. | day\_type | Blocks for day\_of\_week ( dblk1, dblk2, dlk3, dblk4 ) |

The blocks in point 6, 7, 8 above are calculated using quartiles. For example, to calculate hour\_type blocks, the following steps were done:

1. The mean demand for each hour ( 0-23 ) was calculated for the entire dataset.

2. The 1, 2, 3 quartile was calculated for all mean demand values.

3. Based quartile values, the block values were assigned to an hour value ( blk1: mean\_demand <= q1, blk2 : q1 < mean\_demand <= q2, blk3: q2 < mean\_demand <= q3, blk4: mean\_demand > q3 )

The same technique was applied to calculate month\_type and day\_type blocks.