Assumptions:

- Missing cells in the spreadsheet represent that there were no repairs for the part number during that particular week and will be filled in with a value of 0.
- For the purpose of creating a datetime object in Python from Week and Year, I will assume that Monday is the start of the week (i.e., bases weeks from Monday).
- Considering that week 0 represents the first week (and possibly a partial week) of the year, I will drop this week because it may contain incomplete data for the week.

General observations:

Almost all parts were not used in repairs in week 19 of 2015.
Parts were not used consistently in repairs for part number 9 until week 44 of 2015.
No parts were used in repairs for part number 4 beginning week 33 of 2016.
No parts were used in repairs for part 17 until week 9 of 2016.

Summary statistics description:

Overall (2015-2017):
The overall average number of parts per week used in repairs decreased from part number 1 to part number 26 where part 1 average was 934 and part 26 average was 136. The minimum number of parts used was 0 for all part numbers except 21, 22, and 26. The maximum number of parts used was 1689 for part number 5.

2015:
The average number of parts per week used in repairs in 2015 was the greatest for part number 4 (mean = 648). The lowest average number of parts per week used in repairs was for part number 17 (mean = 0) which had no repairs in 2015.

2016:
The average number of parts per week used in repairs in 2016 was the greatest for part number 1 (mean = 990). The lowest average number of parts per week used in repairs was for part number 26 (mean = 86).

2017:
The average number of parts per week used in repairs in 2017 was the greatest for part number 5 (mean = 1488). The lowest average number of parts per week used in repairs was for part number 4 (mean = 0) which had no repairs in 2017.

Trends examining line plots over time:

Parts 1, 2, and 3 had the largest number of parts used in repairs and had a generally steady increase from 2015-2017. There were, however, a number of drops at the end of 2015, just prior to April 2016, and at the end of 2016.

Part 4 had a similar trend to parts 1, 2, and 3. However, part 4 had no repairs beginning in mid-August of 2016 until the end of 2017.

Part 5 had a very small number of repairs in 2015 and until August 2016. The number of repairs significant increased beginning in September 2016 and steadily increased until the end of 2017.

Part 6, 7, and 6 steadily increased from 2015-2017. There were, however, a number of drops at the end of 2015, just prior to April 2016, and at the end of 2016.

Part 9 had a similar trend to parts 6, 7, and 8. However, part 9 had almost no repairs beginning in 2015 until late-October 2015.

Part 10 was relatively constant over the time period but did have a number of spikes over time.

Part 11 was constant through 2015 but then began significantly increased in early 2016. There was a large drop at the end of 2016.

Parts 12, 13, and 14 varied quite significantly over the time period and had a number of drops and peaks at various time points.

Part 15 had a relatively low number of repairs through 2015. The number of repairs significantly increased beginning in 2016 and steadily increased through 2017. There were a number of drops just prior to April 2016, at the end of 2016, and after April 2017.

Part 16 generally had less than 200 repairs per week over the time period. However, there were some significant spikes in December 2016 and February/March 2017.

Part 17 had no repairs through February 2016. Then had a sharp increase in repairs until October 2016. Then had a steady decline in repairs through the end of 2017 with a large drop at the end of 2016.

Part 18 varied significantly over the time period with quite a few peaks and drops. However, it appears to have stabilized beginning in January 2017.

Part 19 had a relatively low number of repairs through 2015. The number of repairs significantly increased beginning in 2016 and steadily increased through 2017. There was a large increase prior to April 2016 and a large drop at the end of 2016.

Part 20 had a generally steady increase through the time period but had a significant increase after January 2017.

Part 21 increased and decreased though the time period with peaks around February and March 2016.

Parts 22 and 26 had a significant decrease in repairs from 2015 through 2017. There were large drops at the end of 2015 and just prior to April 2016.

Parts 23 and 25 had a relatively constant increase over the time period. There were, however, a number of drops at the end of 2015, just prior to April 2016, and at the end of 2016.

Part 24 generally increased through April 2016 and then dropped in repairs until June 2016. Then, there was an increase until the end of 2017 with a few drops at the end of 2016.

Predict next year's parts demand (forecast at part level and weekly level):

In order to make predictions for next year's parts demand, I built a model that will allow us to forecast at both the part and weekly level. The type of model that I chose was an OLS linear regression model because our outcome variable (number of parts) is a continuous measurement. Week number and part number will be our independent variables. For the purpose of modeling, I created dummy variables from week number and part number. I dropped the first dummy variable created for both week number and part number. This resulted in a total of 76 features to be included in the modeling (51 features for week and 25 features for part number).

In order to predict model out-of-sample accuracy, I chose the evaluation procedure train-test-split. I split the data into training and testing datasets and varied the proportion of the dataset to include in the training and testing datasets for comparison. I chose to perform a 70/30 split, 80/20 split, and 90/10 split to evaluate the model performance. I then used both the R-squared and RMSE evaluation metrics to quantify the performance of the model.

The following table shows the results for the evaluation metrics for each of the splits described above.

	R-squared	**RMSE**
70/30	0.517	222
80/20	0.514	228
90/10	0.484	215

I decided that the 70/30 split would be most appropriate because it has the highest R-squared and while not the lowest RMSE, it had a good balance between the other splits.

After selecting the appropriate data split and model as I described above, then I can use the regression coefficients for each of the features (week number and part number) in order to make predictions for next year's parts demand at the week and part number level.

Build a systematic solution to help predict thousands of other parts:

While the value of R-squared for the selected model is admittedly lower than hoped, it is possible to improve the model performance by including additional predictors of the number of parts in the model. I would ask the part demand planner if there is additional information that could be provided so that I could add additional features in the regression model and potentially improve model fit.

This model can also be modified in order to handle additional part numbers. However, this does mean that the model would need to be retrained and re-fit because new features would be added to the model.