Reproducible Supply Chain Component and SKU Risk Evaluation

Jacob Hancock

University of Massachusetts Amherst

Department of Mechanical & Industrial Engineering

MIE 496 - Independent Study

May 12th, 2020

Table of Contents

Overview

Introduction

The basis for this paper is previous work done in Fleury et. al (2019), which describes that manufactured products may be made up of multiple components that may each have different arrival times. The difference in arrival times can often lead to a slower production time in anticipation for missing parts. Many factors can be attributed to this difference, including miscommunication, issues within logistics, quality assurance, or even economic concerns. In turn, costs pertaining to inventory could increase as parts that have already arrived will need to be stored until the missing required parts are received. Longtime trusted suppliers usually are able to account for most lateness, but sometimes delays are beyond the control of that supplier, i.e. through material shortage, acts of God, or failure of tier 2 or tier 3 suppliers that supply parts to a tier 1 supplier. These effects resonate throughout the entire supply chain, with the consequences most felt by the consumer or manufacturer that is unable to receive required goods on time. To counter this, most companies utilize some sort of buffering strategy that accounts for some amount of delay. This buffer entails keeping a defined amount of inventory in stock that would account for some level of delay. Prokle et. al (2019) shows that having a defined ordering strategy results in significant inventory savings in the aerospace industry. The reason for this is because inventory consisting of having all components that wait for one slower arriving part in the system is more frequent than inventory that is made up of ordering each individual part in advance according to its own delay distributions.

Problem

The motivation for this project was brought on by our industrial partner, which is a company that manufactures various assembled consumer goods. Since a manufacturer can sell up to thousands of products that each have components sourced within a supply chain, there are many variables that are factored into whether a part will arrive on time. If a component arrives too early, costs will be incurred to store it. On the other hand, if a part or parts arrive too late, storage costs will once again be incurred for the parts that are already in inventory but also the final assembly may take longer to complete, thus increasing production costs.

Objective

The end goal of this project is to make a Python script that allows for a range of buffering times to be inputted and from those values calculate both component and SKU risk for each value. For component risk, historical delivery data will be parsed and only components who did not arrive on time will be used. SKU risk will be calculated by comparing components with risk of being late to a component to SKU bill of materials (BOM). Only late components will be used to calculate SKU risk because early component arrivals do not cause any sort of SKU risk. These results will then be graphed for comprehensibility in the form of histograms and scatter plots. From this exploration, our industrial partner will be able to determine an optimal buffering value that balances feasibility and risk.

Literature Review

Given geographic dispersion of modern supply chains, there are many risks that could be incurred, leading to slowed down shipping times, delayed production, and high costs. Over time, efficiency has improved and in turn companies now are able to keep smaller inventories. As a result, an amount of rigidity has been placed in supply chains, especially global supply chains. Rigidity will not allow for a supply chain to compensate for major disruption as they will not have the capability to account for inventory changes. It is believed that around 80% of companies are at risk of a major supply chain disruption. Even while corporate leaders support risk-management strategies, it is believed that there are many challenges with implementation and management of these policies. This can be attributed to a lack of clear and concise tools and resources within companies that could be integrated into already existing supply chain management processes (Kumar et al. 2013). On top of this, there is no standard metric of risk defined within supply chain methodology. The most common definition used within industry and academic research is: "supply risk is defined as the probability of an incident associated with inbound supply from individual supplier failures or the supply market occurring, in which its outcomes result in the inability of the purchasing firm to meet customer demand or cause threats to customer life and safety". Companies have been able to determine what risk metrics are most pertinent for them to focus on, but little is described in terms of supply risk with regards to lead time delivery data for assembled systems. Fleury et. al (2019) define both SKU risk and the risk of components they are composed of.

Initial Analysis

Jupyter Notebooks was chosen to code in due to its superior visualization capabilities as well as ease of use for further research when passed to other researchers. To begin data exploration, the provided Excel documents needed to be imported into a data frame first. This was done using the Pandas package "read_excel" function since the document was an ".xlsx" file. Once imported, it was found in the delivery history sheet that from 2017 to 2019, there were 134,371 total deliveries that contained 4,385 unique components from 105 unique vendors that spanned 14 different countries. The second sheet contained the component to SKU BOM. Once all missing values were removed from this sheet, there were a total of 1,218 unique SKUs and 1,518 unique components. This would mean that there could only be a maximum of 1,518 components that could be used to calculate SKU risk from the delivery history sheet.

Component Risk (Example Code Shown in Appendices 1 & 2)

The metric created in Fleury et. al (2019) that determines component risk is defined as the percentage of deliveries for a component that experiences a delay greater than 10 days. In this paper, component risk will be solely defined as the percentage of components whose delay is greater than a user-inputted number of days, to allow for comparison between ranges.

$$\delta = \frac{n}{N}$$

$$\delta = component\ risk$$

Where n = number of deliveries with delay greater than a user-defined number of days

N = total number of deliveries per component

Equation 1: *Component Risk*

As base visualizations, I created histograms for deliveries that were greater than 5 and 10 days overdue and for deliveries that were greater than 5 and 10 days early. These visualizations do not factor in components that have a risk of 0 as most components do not have any risk. The x-axis for each histogram is the calculated component or SKU risk, depending on the chart. These risks range between 0 (0%) to 1.0 (100%) with intervals of 0.1 in between. On the y-axis are the total number of components or SKUs that fall into these risk ranges. For the scatter plots, the x-axis is a range of days used to define lateness or earliness, starting with 1 day, ending with 20 days, and having an interval of 1 day between each point. The y-axis for the scatter plots is the average calculated component risk.

5 Days Late

In Figure 1, most components that contained deliveries that were more than 5 days overdue fell within the range of having between 20% and 40% risk (about 1,500 components), with few having a risk greater than 60% (less than 350 components). That said, about 150 of those components that had a risk greater than 60% had a risk of 100%. Of the 4,385 components total, there are 2,973 components that have a risk greater than 0% (67.8%) and 1,412 components that have a risk of 0% (32.2%).

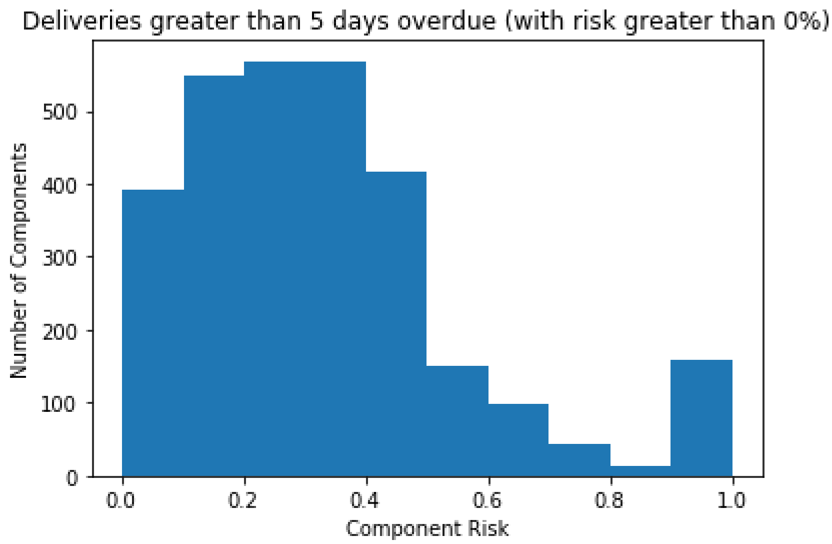


Figure 1: *Breakdown of Component Risk for Deliveries Greater than 5 Days Overdue (67.8% of total components)*

5 Days Early

In Figure 2 on the other hand, of components that had deliveries with risk of being earlier than 5 days, most had a risk less than 20% (about 1,300 components). This indicates that when the metric for risk is set to 5 days early or late for a shipment, there is a higher risk that a component will be late rather than early. Of the 4,385 components total, there are 2,245 components that have a risk greater than 0% (51.2%) and 2,140 components that have a risk of 0% (48.8%).

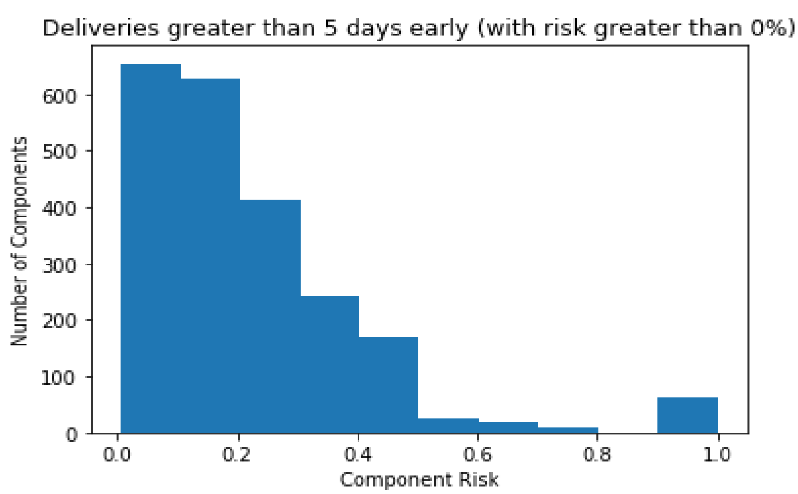


Figure 2: *Breakdown of Component Risk for Deliveries Greater than 5 Days Early (51.2% of total components)*

10 Days Late

In Figure 3, component risk is defined as the number of components that contain shipments that arrived greater than 10 days late. In this range, most components have less than a 20% risk (about 1,500 components). About 600 components have a risk greater than 20%. Once again, there are still some components with a risk of 100% (about 50 in total). Of the 4,385 components total, there are 2,316 components that have a risk greater than 0% (52.8%) and 2,069 components that have a risk of 0% (47.2%).

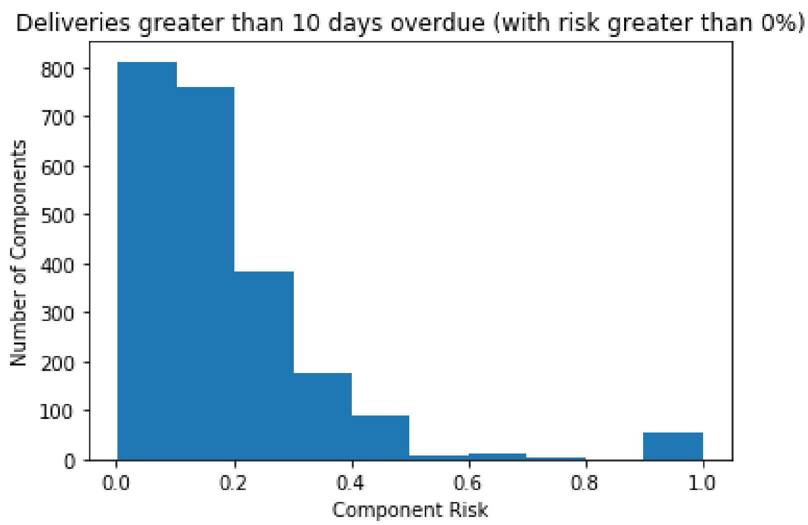


Figure 3: *Breakdown of Component Risk for Deliveries Greater than 10 Days Overdue (52.8% of total components)*

10 Days Early

For Figure 4, component risk is defined as components that contained deliveries that were more than 10 days early. These components mostly had a risk of less than 10% (about 850 components). This is a significant drop in comparison to the average risk of a component arriving at least 10 days late (around 20%). Of the 4,385 components total, there are 1,230 components that have a risk greater than 0% (28%) and 3,155 components that have a risk of 0% (72%).

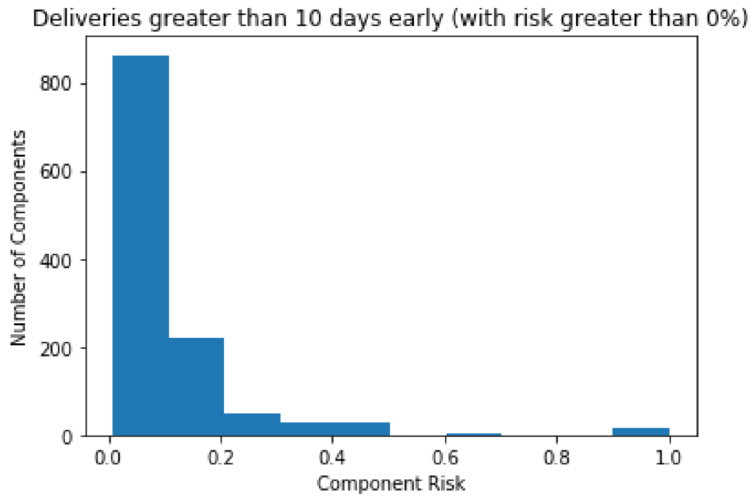


Figure 4: *Breakdown of Component Risk for Deliveries Greater than 10 Days Early (28% of total components)*

Average Risk of Lateness Over a Range of Days

Figure 5 utilizes a range of 1 to 20 days as the number of days late required for a delivery to be considered late. For each day, the average component risk is calculated and plotted. In this range, an almost linearly decreasing trend can be followed. At 1 day, the average component risk is about 50% and at around 17 days component risk almost completely levels out at around 7%.

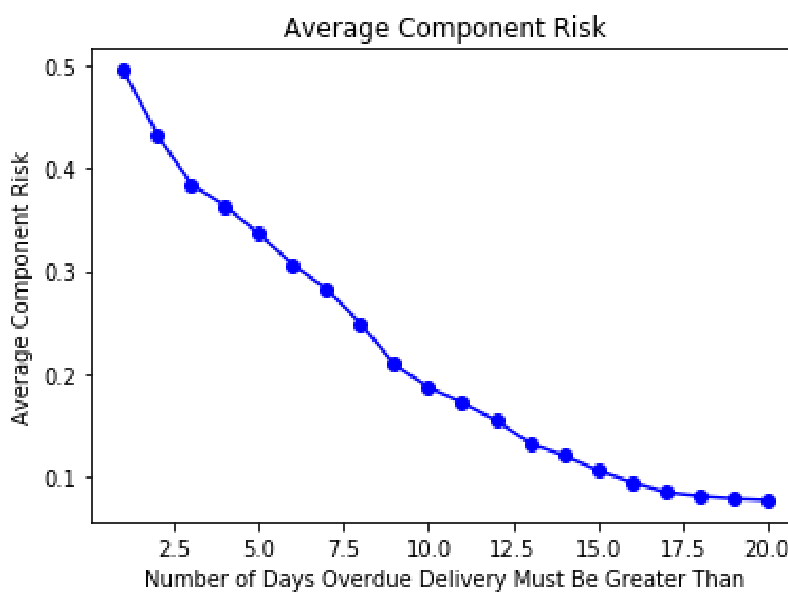


Figure 5: *Average Component Risk for Deliveries Greater than 1 to 20 Days Overdue*

Average Risk of Earliness Over a Range of Days

Figure 6 on the next page also uses a range of 1 to 20 days as the number of days early a shipment must be early by to have risk. Once plotted, an exponentially decreasing trend can be followed. Initial component risk for day 1 is at about 40%, which is also when it is at its highest. At around 10 days, component risk tapers out to about 10%. After that, risk maintains a level of about 10% until the maximum of 20 days is reached.

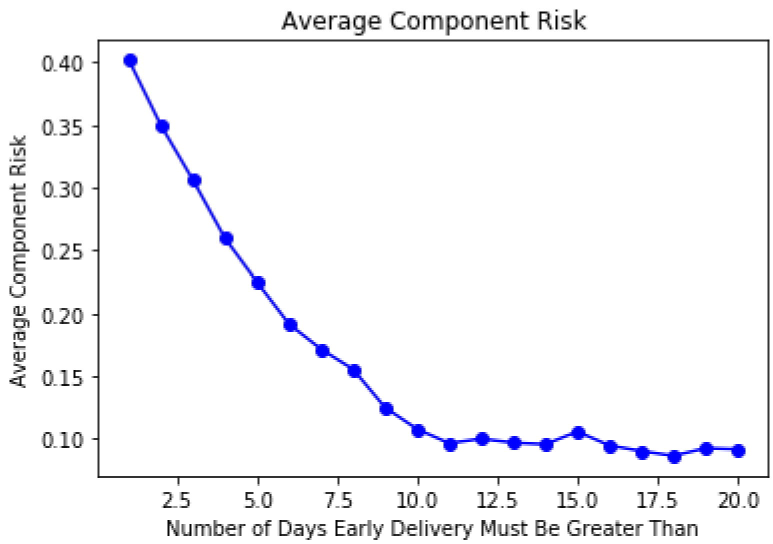


Figure 6: *Average Component Risk for Deliveries Greater than 1 to 20 Days Early*

SKU Risk (Example Code Shown in Appendices 3 & 4)

Before any calculations can be made, the component values in the sheet containing SKU-component combinations need to be converted to match the data structure of components in the delivery history sheet. First, these values are converted to integers to make each value a whole number, then converted to a string value. Once completed, a data frame is created that contains overlapping component ID's contained in the Bill of Materials and components delivered from the delivery history sheet. Once compiled, SKU risk is calculated using the formula contained in Fleury et. al (2019). Since early arrivals do not pose SKU risk, they are left out of these calculations.

$$SKU\ RISK = 1 - \Pi_{i=1...n}(1 - \delta_i)$$

$$where\ \delta_i = component\ i\ risk$$

Equation 2: *SKU Risk*

5 Days Late

Initially, lateness is defined as a delivery arriving more than 5 days late. For components that meet that requirement, the risk for the corresponding SKUs that they make up is calculated. Just as for component risk calculations previously, components with risks of 0 are left out. Of the 1,218 total SKUs, there are 1,210 SKUs that have a risk greater than 0% (99%) and 8 SKUs that have a risk of 0% (1%). Surprisingly, many SKUs that have any risk have over a 90% risk of being late (more than 500 out of about 1,100 total). This is plotted in Figure 6 below.

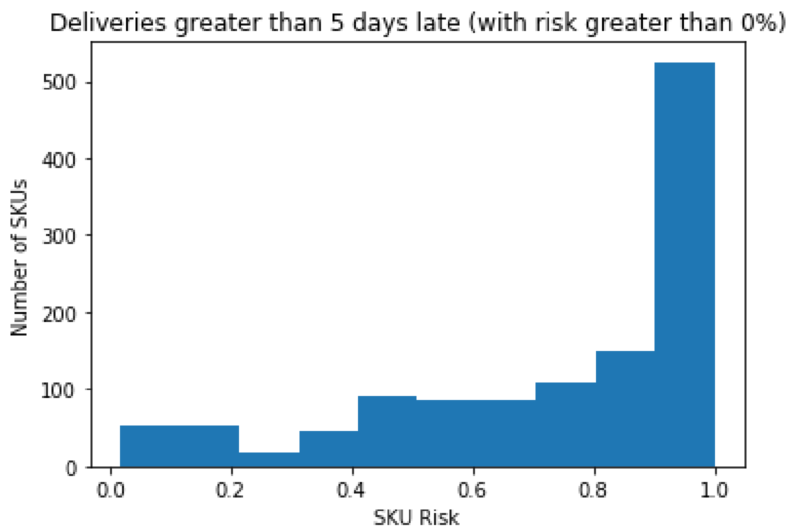


Figure 6: *Breakdown of SKU Risk for Deliveries Greater than 5 Days Late*

10 Days Late

When late deliveries are defined as being more than 10 days late, there is a much more even distribution as compared to when 5 days late is the defined metric for lateness. Of the 1,218 total SKUs, there are 1,168 SKUs that have a risk greater than 0% (95.9%) and 50 SKUs that have a risk of 0% (4.1%). The average risk for SKUs using this metric is calculated to be 55%

as opposed to the previously calculated 80% for when arriving greater than 5 days late is considered late. This distribution is plotted below in Figure 7.

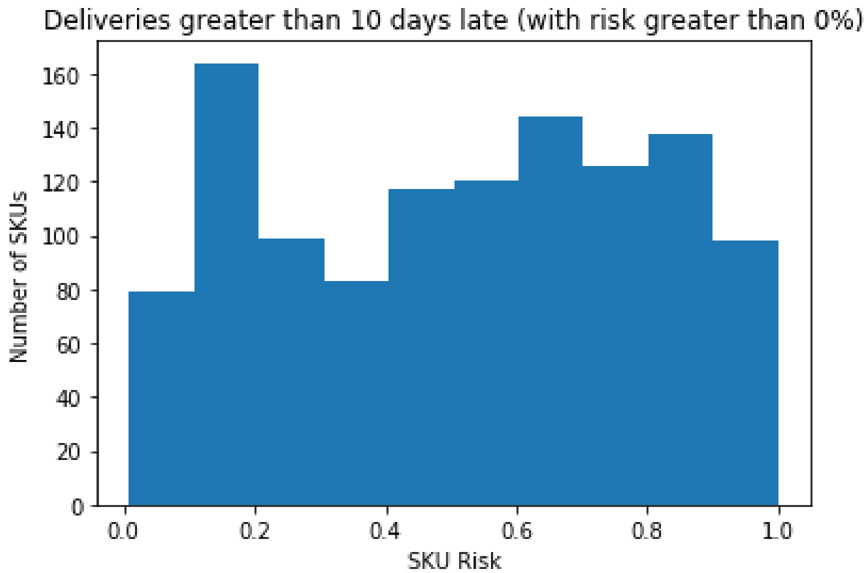


Figure 7: *Breakdown of SKU Risk for Deliveries Greater than 10 Days Late*

Average Risk of Lateness Over a Range of Days

As with average component risk calculation, a range of 1 to 20 days is utilized as the number days required for a delivery to be considered late for SKU risk. Just as with component risk, a linearly decreasing trend can be followed. Initially at day 1 the average SKU risk is at its highest (about 95%) and at 20 days the lowest amount of SKU risk is reached (about 5%). Figure 8 below is the plot for this distribution.

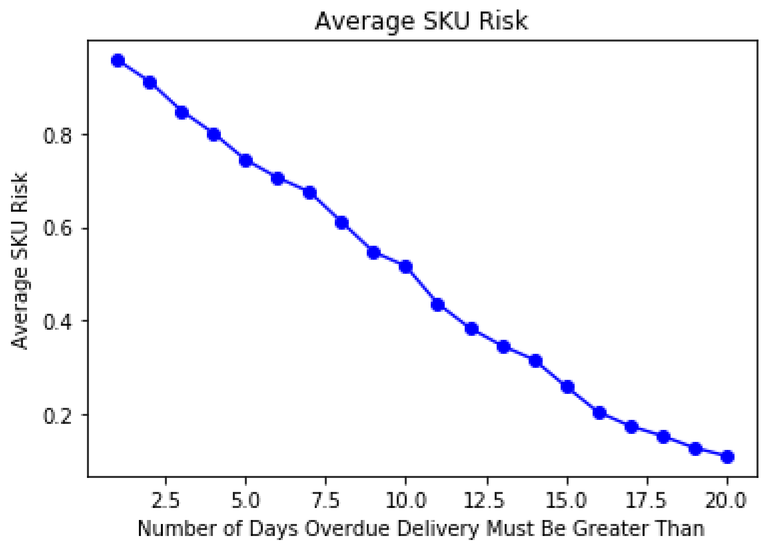


Figure 8: *Average SKU Risk for Deliveries Greater than 1 to 20 Days Late*

Conclusion

After calculating the average component risk for early arriving and late components separately over the course of the same range of days, it can be determined that more components are likely to be late than early. The plot for when a component is late by a range of days decreases almost linearly and flattens out around 17 days while the risk of when a shipment arrives early decreases exponentially and tapers out around 10 days. In the distribution for 10 days, there are still several components that have a risk of 100% (51 in total). Component risk has a compounding effect on SKU risk, leading to a much higher risk of production delay for assembly. For those 51 components that have 100% risk of being greater than 10 days late, perhaps it would be good to examine if there is any overlap between vendor or country of origin or the total amount of shipments they were a part of throughout 2017-2019. These components may only have been delivered once or twice and it may have been a fluke that they had been delivered late. That said, when it comes to an optimal allowance for early shipments, a similar route of examination should be undertaken as there are about 20 components that have 100% risk of being early. Once our industrial partner feels satisfied with their analysis of these components, a next step they could take is developing a buffering strategy that incorporates an optimal allowance for the number of days a shipment can arrive in time to be considered satisfactory.

SKU risk calculation is based on a threshold delay duration. As the number of days increases for threshold delay classification, the amount of risk decreases. The highest average SKU risk is incurred when lateness is classified as arriving more than one day late (about 95% risk). The lowest average SKU risk is incurred when lateness is classified as any shipment arriving more than 20 days late (about 5% risk). This further supports the original suggestion of focusing on determining the optimal number of days late a shipment must be to be considered late as component risk will have a strong effect on subsequent SKU risk. Analysis and risk tolerance will vary between industries and so data exploration and concluding decisions will vary. For example, some companies with not as frequent shipments that carry costly components will potentially require a lower tolerance to maintain low inventory and production costs. Companies with a higher volume of deliveries containing more inexpensive parts on the other hand, may be able to maintain a higher risk tolerance as it could be inexpensive to keep a larger inventory that compensates for late deliveries. As stated before, these decisions will vary from company to company and will need to be explored individually by that company to determine a metric that works best for them.

Further Research

While there are limitations to these calculations due to the rudimentary analysis of components and SKUs, the breadth of this work can be improved upon. First, including both vendor and country of origin into risk calculations could add a fair amount of value. Doing this would allow for our industrial partner to be able to potentially find trends of underperforming vendors or countries of higher risk. The same component may be delivered from multiple sources and this more complex breakdown will allow for a comparison of risks between different vendors of the same component. The result of this could be that certain vendors may have a trend of delivering on time while another may oftentimes be late.

Another potential for improvement could lie in adding weights to components or SKUs of higher value. More weight could be added to higher value items, leading to a greater importance in calculated risk for these items as they will have a more detrimental effect on our industrial partner dependent on incorrect arrival times.

Next, if another sheet were to be provided that includes more SKU component combinations that make up for the 2,867 missing components, more accurate results could be calculated. This is dependent on the number of deliveries that these missing components are in. If each missing component were delivered only once, that would account for only 2,867 shipments out of the total 134,371 deliveries, or about 2%. There is also the possibility that many of those components make up a large enough percentage of the total, which would lead to a difference in risk calculations for both component and SKU risk, respectively.

In the SKU risk calculation, the risk of early arrivals is left out as that would not impact SKU risk. Even so, developing a formula to calculate early arrival risk and how it impacts SKUs could be helpful in managing a buffer inventory. Having the minimal amount of stock needed to manufacture will minimize holding costs.

References

Fleury P, Muriel A. (2019) Assessing Supplier Delivery Risk. University of Massachusetts Amherst.

Prokle M, Beladi F, Muriel A, Subbu R, Greene C, Heslin J. (2019). Supply chain Synchronization for Assembled Systems with Long and Highly Variable Component Lead Times. University of Massachusetts Amherst.

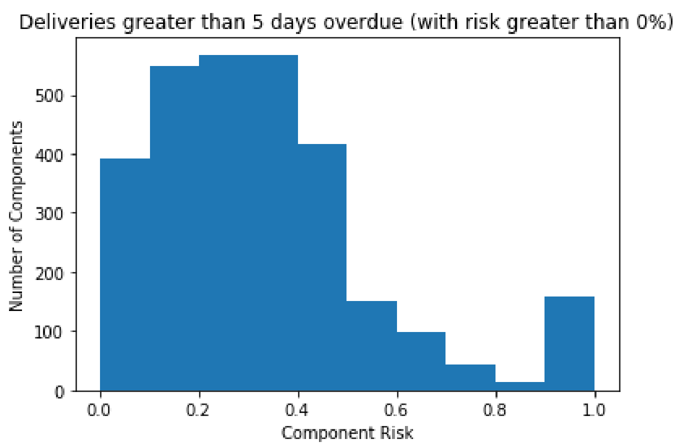
Kumar S, Boice B, Shepherd M. (2013). Risk Assessment and Operational Approaches to Manage Risk in Global Supply Chains. *Transportation Journal, 52*(3), 391-411. doi:10.5325/transportationj.52.3.0391

Appendix

Appendix 1: *Example Code for Component Risk Calculation and plotting for Deliveries More than 5 Days Overdue*

In [142]:
```
deliveries = Lead_time_performance.groupby(['Material'])['PDT Delta'].apply(lambda x: (x > 5).sum())
Component_risk = np.divide(deliveries, total_deliveries)
Component = [x for x in Component_risk if x != 0.0 and x != 'nan']
plt.hist(Component)
plt.title('Deliveries greater than 5 days overdue (with risk greater than 0%)')
plt.ylabel('Number of Components')
plt.xlabel('Component Risk')
```

Out[142]:
```
Text(0.5, 0, 'Component Risk')
```

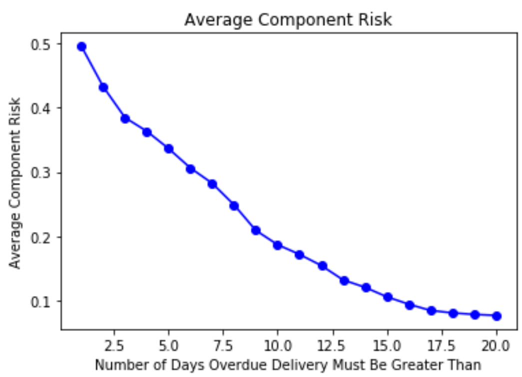


Appendix 2: *Example Code for Average Component Risk Calculation and plotting for Deliveries Ranging from More than 1 Day to More than 20 Days Overdue*

Average risk from greater than 1 to 20 days overdue plotted (with risk greater than 0%):

In [9]:
```
days = [*range(1, 21, 1)]
Average_risk = []
for day in days:
    deliveries = Lead_time_performance.groupby(['Material'])['PDT Delta'].apply(lambda x: (x > day).sum())
    Component_risk = np.divide(deliveries, total_deliveries)
    Component = [x for x in Component_risk if x != 0.0]
    Average_risk.append(np.divide(np.nansum(Component), len(Component) - len([0 for x in Component if math.isnan(x)])))
plt.plot(days, Average_risk, '-bo')
plt.title('Average Component Risk')
plt.xlabel('Number of Days Overdue Delivery Must Be Greater Than')
plt.ylabel('Average Component Risk')
```

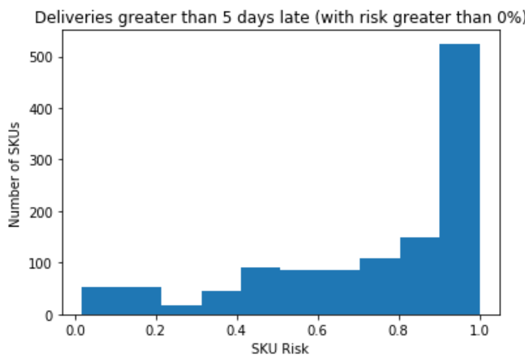
Out[9]:
```
Text(0, 0.5, 'Average Component Risk')
```



Appendix 3: *Example Code for SKU Risk Calculation and plotting for Deliveries More than 5 Days Overdue*

In [143]:
```
deliveries = Lead_time_performance.groupby(['Material'])['PDT Delta'].apply(lambda x: (x > 5).sum())
Component_risk = np.divide(deliveries, total_deliveries)
Forecast_material = SKU_mapping.loc[SKU_mapping['Component'].isin(SKU_mapping['Component'])]
Forecast_material['Risk'] = Forecast_material['Component']
Forecast_material = Forecast_material.loc[Forecast_material['Risk'].isin(Component_risk.index)]
Risk = Component_risk.to_dict()
Forecast_material = Forecast_material.replace({"Risk": Risk})
Forecast_material = Forecast_material[Forecast_material['Risk'] != 0]
SKU_Risk = Forecast_material.groupby(['Forecast Material'])['Risk'].apply(lambda x: np.subtract(1, np.prod(np.subtract(1, x))))
plt.hist(SKU_Risk)
plt.title('Deliveries greater than 5 days late (with risk greater than 0%)')
plt.ylabel('Number of SKUs')
plt.xlabel('SKU Risk')
```

Out[143]:
```
Text(0.5, 0, 'SKU Risk')
```



Appendix 4: *Example Code for Average Component Risk Calculation and plotting for Deliveries Ranging from More than 1 Day to More than 20 Days Overdue*

In [119]:
```
days = [*range(1, 21, 1)]
Average_risk = []
for day in days:
    deliveries = Lead_time_performance.groupby(['Material'])['PDT Delta'].apply(lambda x: (x > day).sum())
    Component_risk = np.divide(deliveries, total_deliveries)
    Forecast_material = SKU_mapping.loc[SKU_mapping['Component'].isin(SKU_mapping['Component'])]
    Forecast_material['Risk'] = Forecast_material['Component']
    Forecast_material = Forecast_material.loc[Forecast_material['Risk'].isin(Component_risk.index)]
    Risk = Component_risk.to_dict()
    Forecast_material = Forecast_material.replace({"Risk": Risk})
    Forecast_material = Forecast_material[Forecast_material['Risk'] != 0]
    SKU_Risk = Forecast_material.groupby(['Forecast Material'])['Risk'].apply(lambda x: np.subtract(1, np.prod(np.subtract(1, x))
    Average_risk.append(np.divide(np.nansum(SKU_Risk), len(SKU_Risk) - len([0 for x in SKU_Risk if math.isnan(x)])))
plt.plot(days, Average_risk, '-bo')
plt.title('Average SKU Risk')
plt.xlabel('Number of Days Overdue Delivery Must Be Greater Than')
plt.ylabel('Average SKU Risk')
```

Out[119]:
```
Text(0, 0.5, 'Average SKU Risk')
```