MSCI 719: Operations Analytics

Assignment 3: Supply Chain Analytics to Manage Blood at VHS Blood Bank

Student: Advait Shah

Faculty: Engineering

Department: Management Sciences

Instructor: Prof. Hossein Abouee Mehrizi

Contents

1. Bags Containers

VHS is considering ordering some containers which make the transportation of the blood bags collected from donors more efficient. These non-reusable containers will be delivered to VHS on a 5-days basis starting from October 1st, 2018. There are various types of containers and VHS tends to order a specific type that is suitable for the bags used for “Whole Blood”, “Platelets”, and “Plasma”. Due to the high inventory cost of the containers, VHS prefers to keep a low inventory of them. Each container has the capacity for three bags. Athmanathan is thinking of creating a forecasting model to determine the number of containers required for the first three periods (October 1st - October 15th, 2018).

1.1. Visualize the supply of each blood product as a time series properly

Here is the head of our dataset:

Fiancial Year	Bag Number	Status Description	Blood Group	RH	Blood Classification	Stock Creation Date	Stock Issue Date	Expiry Date
2017-2018	112PR	Issued	A	POSITIVE	PACKED RED CELLS	01-01-2018	03-01-2018	09-02-2018
2017-2018	118PR	Issued	A	POSITIVE	PACKED RED CELLS	31-12-2017	31-12-2017	29-01-2018
2017-2018	163PR	Issued	O	POSITIVE	PACKED RED CELLS	27-03-2018	31-03-2018	07-05-2018
2017-2018	7728PL	Issued	B	POSITIVE	PLASMA	02-10-2017	22-10-2017	02-10-2018
2017-2018	7728CY	Expired	B	POSITIVE	CRYO	02-10-2017	NA	02-10-2018
2017-2018	7730PL	Issued	B	POSITIVE	PLASMA	02-10-2017	22-10-2017	02-10-2018
2017-2018	7730CY	Expired	B	POSITIVE	CRYO	02-10-2017	NA	02-10-2018
2017-2018	7731CY	Expired	O	POSITIVE	CRYO	02-10-2017	NA	02-10-2018
2017-2018	7731PL	Issued	O	POSITIVE	PLASMA	02-10-2017	22-10-2017	02-10-2018
2017-2018	7736PL	Issued	O	POSITIVE	PLASMA	02-10-2017	22-10-2017	02-10-2018

Figure: Head of Dataset

With this dataset we can understand the supply and demand of each blood products at VHS Bloodbank over one year time period. Stock creation Date gives idea of Bags supplied to VHS, whereas, Stock issue date gives idea of Bags demanded from VHS.

As we want to predict bags supply containers and therefore number of bags supplied for 5 days period for each blood product, we can create pivot table with following head:

Count of Bag Number	Blood Products						
Stock Creation Period	CRYO	FRESH FROZEN PLASMA	PACKED RED CELLS	PLASMA	PLATELETS	SINGLE DONOR PLATELET	WHOLE BLOOD
02-10-2017 - 06-10-2017	36	54	34	66	182	0	24
07-10-2017 - 11-10-2017	13	6	6	17	148	0	18
12-10-2017 - 16-10-2017	25	65	31	25	134	3	16
17-10-2017 - 21-10-2017	3	3	2	3	12	5	7
22-10-2017 - 26-10-2017	23	35	13	37	100	2	18
27-10-2017 - 31-10-2017	7	30	12	28	48	2	16
01-11-2017 - 05-11-2017	24	58	65	25	57	6	16
06-11-2017 - 10-11-2017	1	148	183	1	94	7	12
11-11-2017 - 15-11-2017	47	254	455	57	133	2	23
16-11-2017 - 20-11-2017	6	135	205	7	90	0	7
21-11-2017 - 25-11-2017	1	238	399	2	115	1	4
26-11-2017 - 30-11-2017	4	164	196	5	116	3	19

Figure: Head of Pivot table

On plotting this data, we can create following visualization:

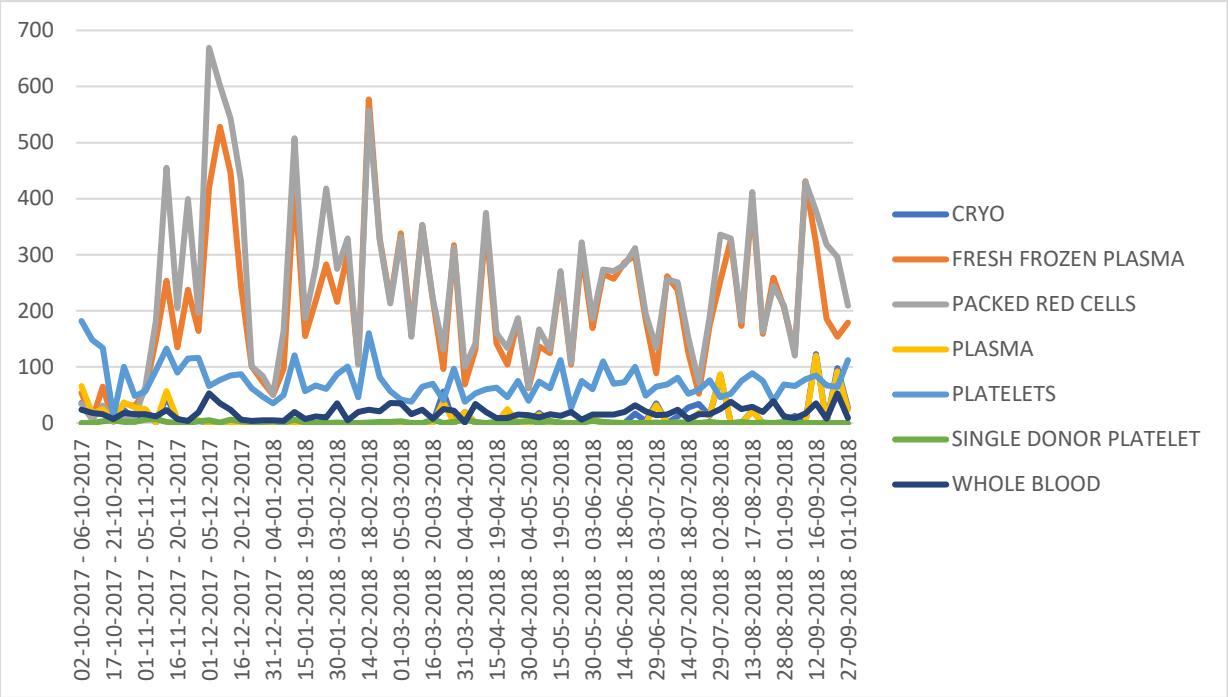

Fig: Visualization of the supply of each blood product as a time series

Now, as given in the question, we are interested in forecasting bags supply and containers mainly for for “Whole Blood”, “Platelets”, and “Plasma”. So, create a separate visualization for these products:

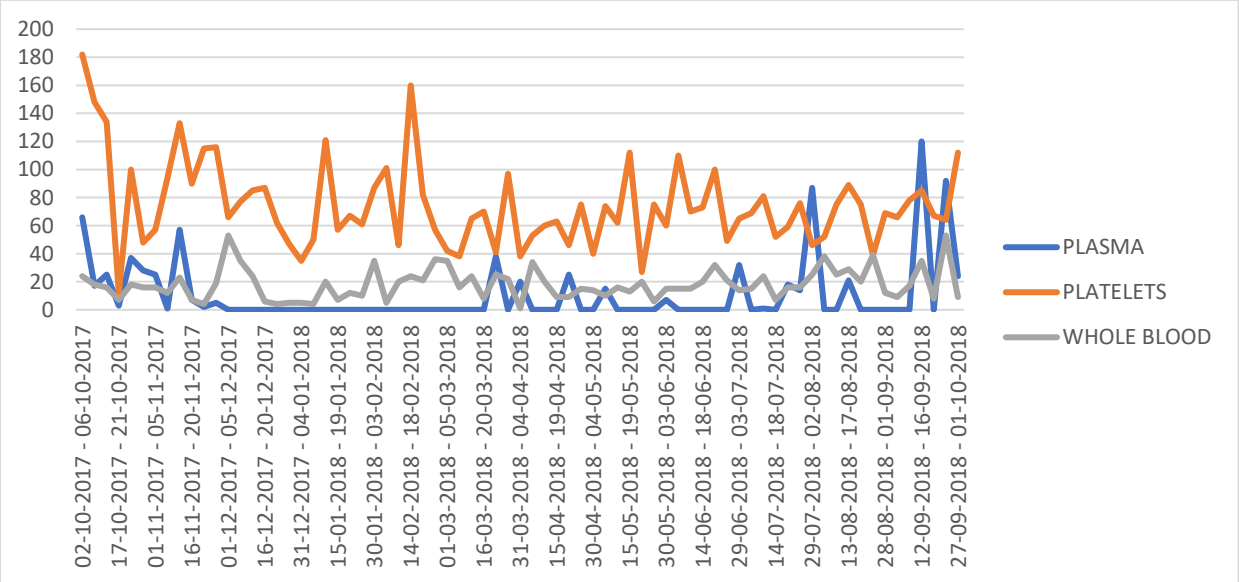

Fig: Visualization of the supply of required 3 blood products as a time series

1.2. Check for the outliers statistically, and if there exists any, remove them from your dataset:

Based on above supply count table, we can detect outliers using Excel whisker plot, which basically shows outliers which are beyond the (Q1-1.5*IQR,Q3+1.5*IQR) range.

Note that Q1 and Q3 here suggests 1st and 3rd Quartile respectively, and IQR is interquartile range i.e. Q3-Q1.

So, we get following plot for outliers detection:

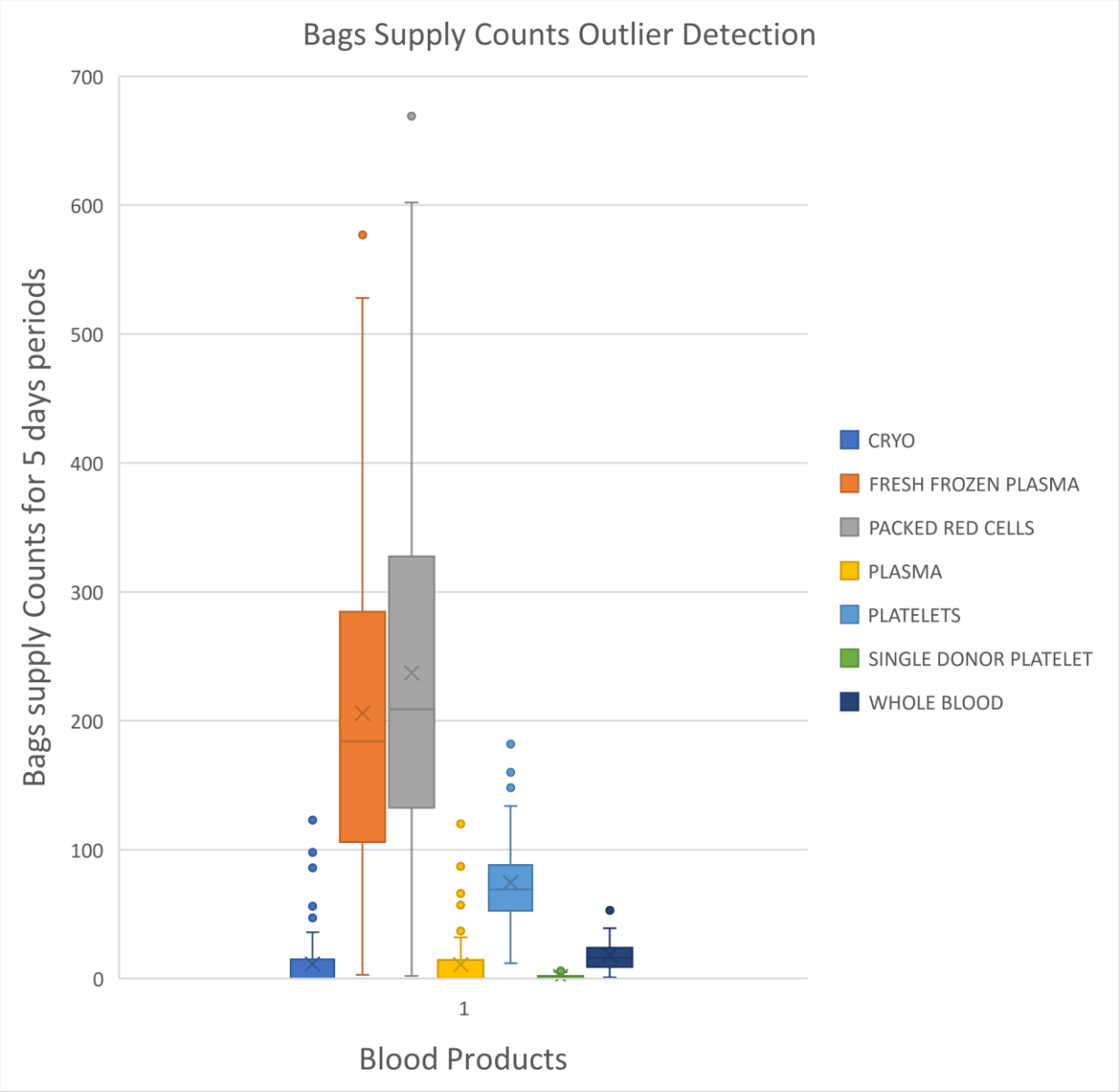

Figure: Outliers detection in blood products supply over 5 days periods

From the plot, we see that there are number of outliers in each of the blood products. As we will be creating forecasting model for one particular product in the upcoming sections, we remove outliers particularly for “Platelets” product based on above mentioned formula using Excel.

After removal of these outliers in the Platelets blood product Bags Supply counts, we get following timeseries chart:

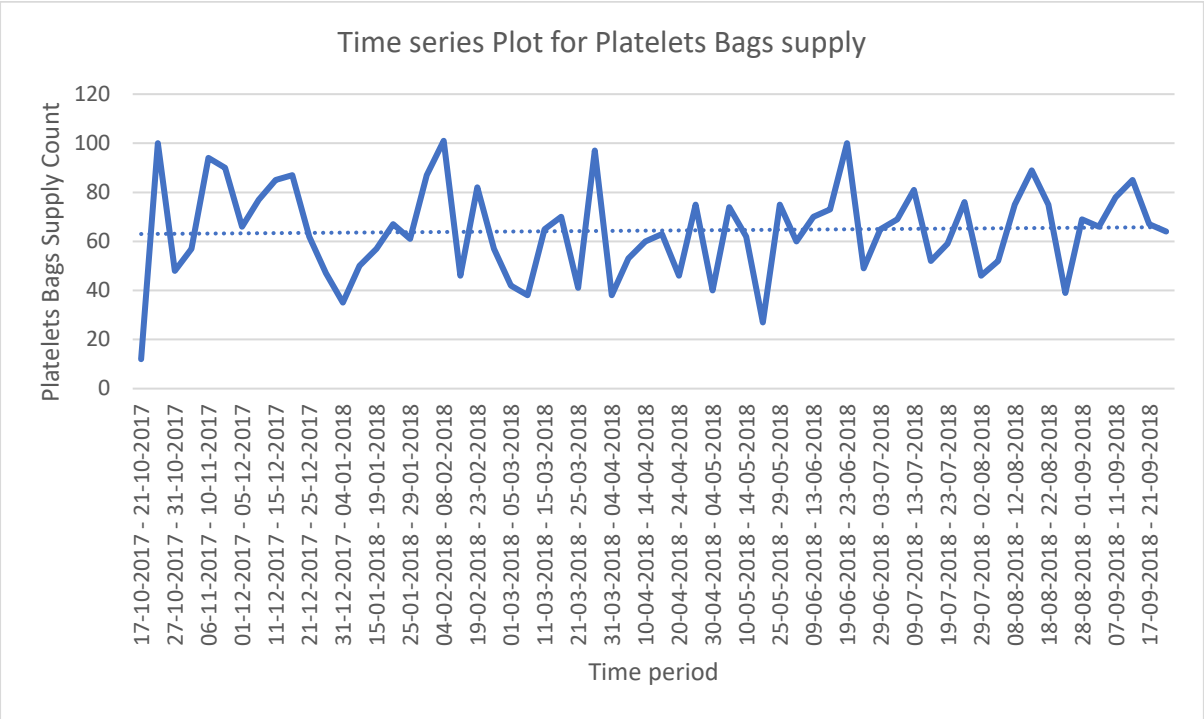

Figure: Time series Plot for Platelets Bags supply Counts

1.3. Do you see any trend in the data? If there is, remove its effect from the time series using a proper method.

From the above time-series plot, we can visually discern that there is no major trend of simple linear or quadratic polynomial degree, as our best-fit dotted-trend-line in the chart is almost horizontal. Had there been any trend, we could have detrend our data by fitting suitable model and calculating residuals, or another method to create detrend data could have been of doing subtraction between two consecutive datapoints.

1.4. Select one blood type and use historical data from July 1st through September 21st to predict the supply from September 1st through September 30th based on the moving average approach. Use different error measures to find the best value of n for the moving average prediction model.

As we have created time series plot for “Platelets” blood product above, which is free from outliers and trend, is perfect to go ahead with forecasting model preparation.

In the Moving Average (MA) approach, to forecast bags counts for the next period, we use following formula,

$$F_{t+1} = \frac{1}{N} \sum_{k=t+1-N}^{t} Y_k$$

Where, F_{t+1}: Forecasted bags supply at time t+1 and Y_k: Actual Supply at time k

So, based on this approach, we could predict bags supply for next periods. And then to evaluate our prediction, we will use following error measures:

RMSE: the Root Mean Squared Error between forecast and actual: $\sqrt{\frac{\sum_{t=1}^{n}(Y_t - F_t)^2}{n}}$

MAPE: the Mean Absolute Percent Error between forecast and actual: $\frac{1}{n}\sum_{t=1}^{n}|Y_t - F_t|/Y_t$

Then, we can train model based on different values of N (i.e. number of prior periods used for moving average calculation and forecast) and measure these errors and select best N, hyper parameter, which minimizes our RMSE error loss function.

After trying with N=3 to N=9, we observe that with N=6, we get the following minimum forecasting error results among all:

Time Period	Actual Bags	Forecast Bags	Sq. error	% error
28-08-2018 - 01-09-2018	69	62.667	40.11111	9.178744
02-09-2018 - 06-09-2018	66	66.5	0.25	0.757576
07-09-2018 - 11-09-2018	78	68.833	84.02778	11.75214
12-09-2018 - 16-09-2018	85	69.3333	245.4444	18.43137
17-09-2018 - 21-09-2018	67	68.6667	2.777778	2.487562
22-09-2018 - 26-09-2018	64	67.3333	11.11111	5.208333

RMSE = 7.99, MAPE = 7.97

1.5. For your chosen time series in the previous part with the same test and training data sets, create exponential models for different values of α. Use different error measures to find the best value of α. Between your best MA and SES, which one performs better?

In the Simple Exponential Smoothing (SES) approach, to forecast bags counts for the next period, we use following formula,

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t$$

Where, F_{t+1}: Forecasted bags supply at time t+1

Y_t: Actual Supply at time t

Ft: Forecasted bags supply at time t

Here also to have accurate forecasting, we need to use optimal α value, which can be determined iteratively by minimizing following RMSE Loss function:

$$Min \sqrt{\frac{1}{n}\sum_{t}(Y - F_t)^2}$$

$$S.t: \quad 0 < \alpha < 1$$

After checking errors for all α values from 0.1 to 0.9 in steps of 0.1, we saw that $\alpha = 0.2$ gave us minimum supply of bags prediction error values among all different α values-based models, as follows:

RMSE = 9.64 and MAPE = 10.44

Based on this comparison of Moving average (MA) and simple exponential smoothing (SES) models optimized for least RMSE, we can say that Moving average (MA) model performs much better in predicting supplies for September month compared to its counterpart, as RMSE with MA is just 7.99, whereas it is 9.64 with SES model.

2. Analytical Report

In the meetings, some managers refer to some statistics which usually arise from their own perception, rather than the actual data that is reflective of the actual status quo. Athamanthan, wants to prepare a comprehensive report, including descriptive charts, graphs, tables, etc. which will help him present his ideas based on precise statistics. Assume that Athamanathan has asked you to prepare such a report based on the available data set.

2.1. List the key statistics you think Athamanathan should focus on (Such as the number of donated bags or ratio of wastage) and visualize them properly. Give brief reasoning for your choices.

[A] We will first focus on the bags wastage in different months of the year.

Total monthly bags received, issued and expired at VHS Blood bank:

Month	Bags Received	Bags Issued	Bags Expired	% Expired
Oct	1309	792	423	32.31
Nov	3385	2176	983	29.04
Dec	4813	4415	339	7.04
Jan	3325	3106	202	6.08
Feb	4204	3870	297	7.06
Mar	3620	3371	241	6.66
Apr	2630	2467	117	4.45
May	2650	2407	188	7.09
Jun	3697	3409	225	6.09
Jul	2769	2444	142	5.13
Aug	4013	3658	236	5.88
Sep	4418	3134	390	8.83

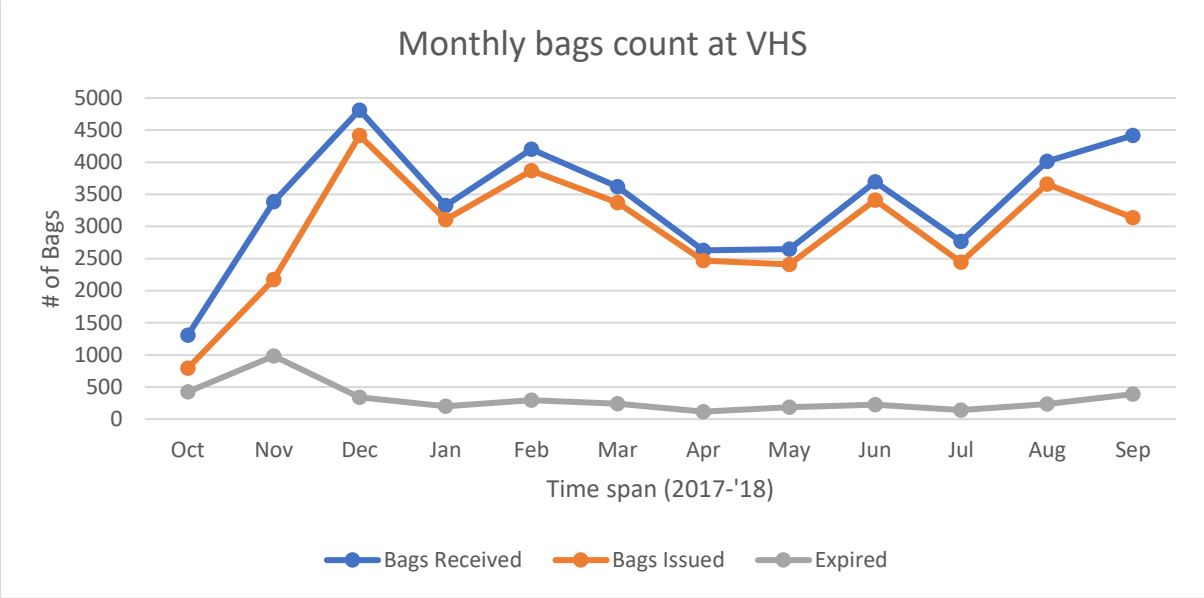

Figure: Monthly no. of bags Supplied, Issued and Expired at VHS

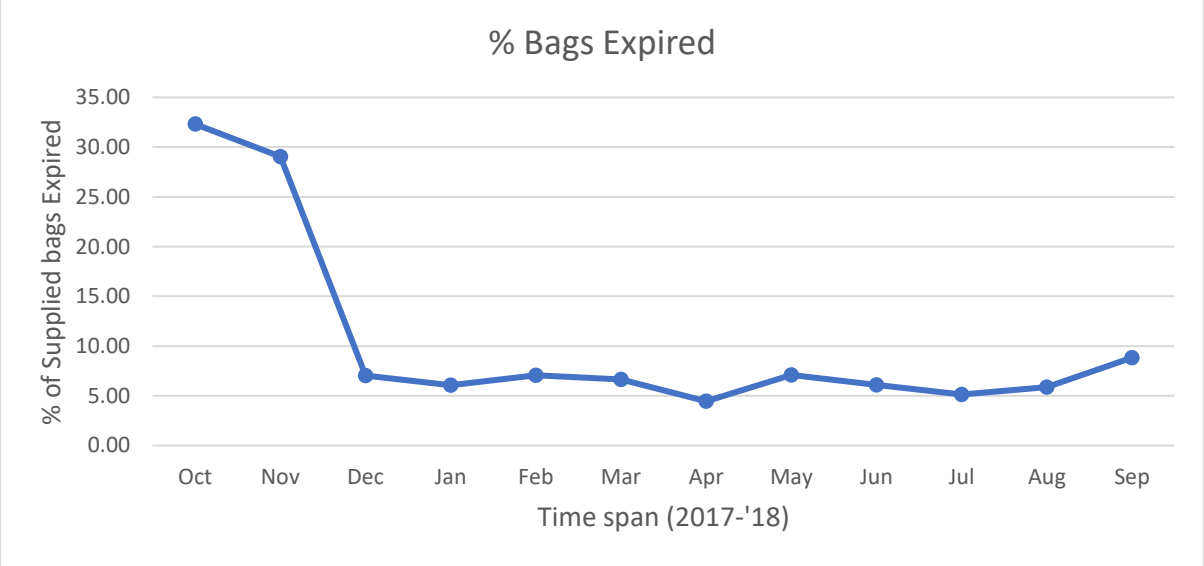

Figure: % of Supplied Bags Expired or Wastage over different months

This plot suggests that in the month of October 2017, November 2017 and September 2018, there is a huge difference between supply and demand, and that explains why there is higher percentage of bags expired in these months. So, a better balance between supply and demand could reduce bags wastage through expiry.

[B] Next, to gain insights into blood product-wise supply, demand, wastage and availability over the months, we can create following visualizations:

Table: Blood Products wise Supply, Demand, Bags available and Expiry for entire year:

Blood Products	Bags Supplied	Issued (Demand)	Expired	Available	% Expired
CRYO	828	301	27	495	3.260869565
FRESH FROZEN PLASMA	15026	14477	18	437	0.11979236
PACKED RED CELLS	17307	15683	1398	101	8.07765644
PLASMA	787	356	0	424	0
PLATELETS	5436	3309	2087	0	38.39220015
SINGLE DONOR PLATELET	117	76	41	0	35.04273504
WHOLE BLOOD	1332	1047	212	35	15.91591592
Aggregate	**40833**	**35249**	**3783**	**1492**	**9.264565425**

Table: Blood Products wise Supply, Demand, Bags available and Expiry for different months:

Product Supply and utilization %ges	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
CRYO	**8.17%**	**2.45%**	**0.00%**	**0.00%**	**0.00%**	**1.55%**	**1.60%**	**1.06%**	**1.19%**	**5.71%**	**1.20%**	**5.93%**
Available	55.14%	24.10%	0.00%	0.00%	0.00%	0.00%	28.57%	60.71%	70.45%	47.47%	81.25%	92.37%
Expired	20.56%	1.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.53%
Issued	20.56%	73.49%	0.00%	0.00%	0.00%	100.00%	71.43%	39.29%	29.55%	52.53%	18.75%	6.11%
FRESH FROZEN PLASMA	**14.74%**	**29.45%**	**37.67%**	**36.96%**	**41.63%**	**41.52%**	**37.72%**	**38.30%**	**40.01%**	**34.49%**	**40.74%**	**33.16%**
Available	9.84%	16.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%	0.00%	0.00%	0.24%	17.00%
Expired	0.52%	1.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%
Issued	88.08%	80.74%	98.57%	99.43%	99.60%	99.93%	99.80%	99.41%	99.66%	99.90%	98.96%	82.12%
PACKED RED CELLS	**7.49%**	**44.40%**	**50.49%**	**48.36%**	**42.86%**	**41.96%**	**42.02%**	**41.17%**	**42.20%**	**39.44%**	**41.86%**	**41.17%**
Available	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.18%	5.39%
Expired	35.71%	39.12%	6.01%	0.93%	4.50%	7.44%	0.27%	4.22%	0.77%	1.10%	5.30%	14.18%
Issued	64.29%	59.75%	93.00%	98.57%	95.12%	92.43%	99.64%	95.14%	98.59%	98.35%	94.11%	78.45%
PLASMA	**13.45%**	**2.87%**	**0.00%**	**0.00%**	**0.00%**	**1.05%**	**1.71%**	**0.83%**	**0.54%**	**3.79%**	**1.20%**	**5.34%**
Available	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	55.56%	95.45%	80.00%	94.29%	100.00%	91.10%
Expired	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Issued	97.16%	98.97%	0.00%	0.00%	0.00%	100.00%	44.44%	0.00%	20.00%	5.71%	0.00%	8.90%
PLATELETS	**47.67%**	**17.87%**	**8.81%**	**11.76%**	**12.68%**	**9.97%**	**12.97%**	**15.32%**	**12.82%**	**13.00%**	**10.34%**	**11.36%**
Available	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Expired	49.36%	53.39%	41.98%	45.01%	32.46%	27.70%	30.50%	34.24%	44.09%	35.56%	32.77%	22.51%
Issued	50.32%	45.95%	58.02%	54.48%	63.23%	70.91%	69.50%	65.27%	55.91%	64.44%	66.99%	77.29%
SINGLE DONOR PLATELET	**0.92%**	**0.56%**	**0.39%**	**0.33%**	**0.14%**	**0.36%**	**0.42%**	**0.45%**	**0.11%**	**0.18%**	**0.07%**	**0.05%**
Available	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Expired	83.33%	89.47%	26.32%	0.00%	0.00%	0.00%	45.45%	0.00%	0.00%	20.00%	33.33%	100.00%
Issued	16.67%	10.53%	73.68%	100.00%	100.00%	100.00%	54.55%	100.00%	100.00%	80.00%	66.67%	0.00%
WHOLE BLOOD	**7.56%**	**2.39%**	**2.64%**	**2.59%**	**2.69%**	**3.59%**	**3.57%**	**2.87%**	**3.14%**	**3.39%**	**4.59%**	**2.99%**
Available	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	26.52%
Expired	47.47%	49.38%	7.87%	12.79%	38.05%	21.54%	5.32%	3.95%	3.45%	1.06%	5.43%	7.58%
Issued	50.51%	44.44%	85.04%	87.21%	61.95%	78.46%	88.30%	94.74%	95.69%	96.81%	92.39%	59.85%

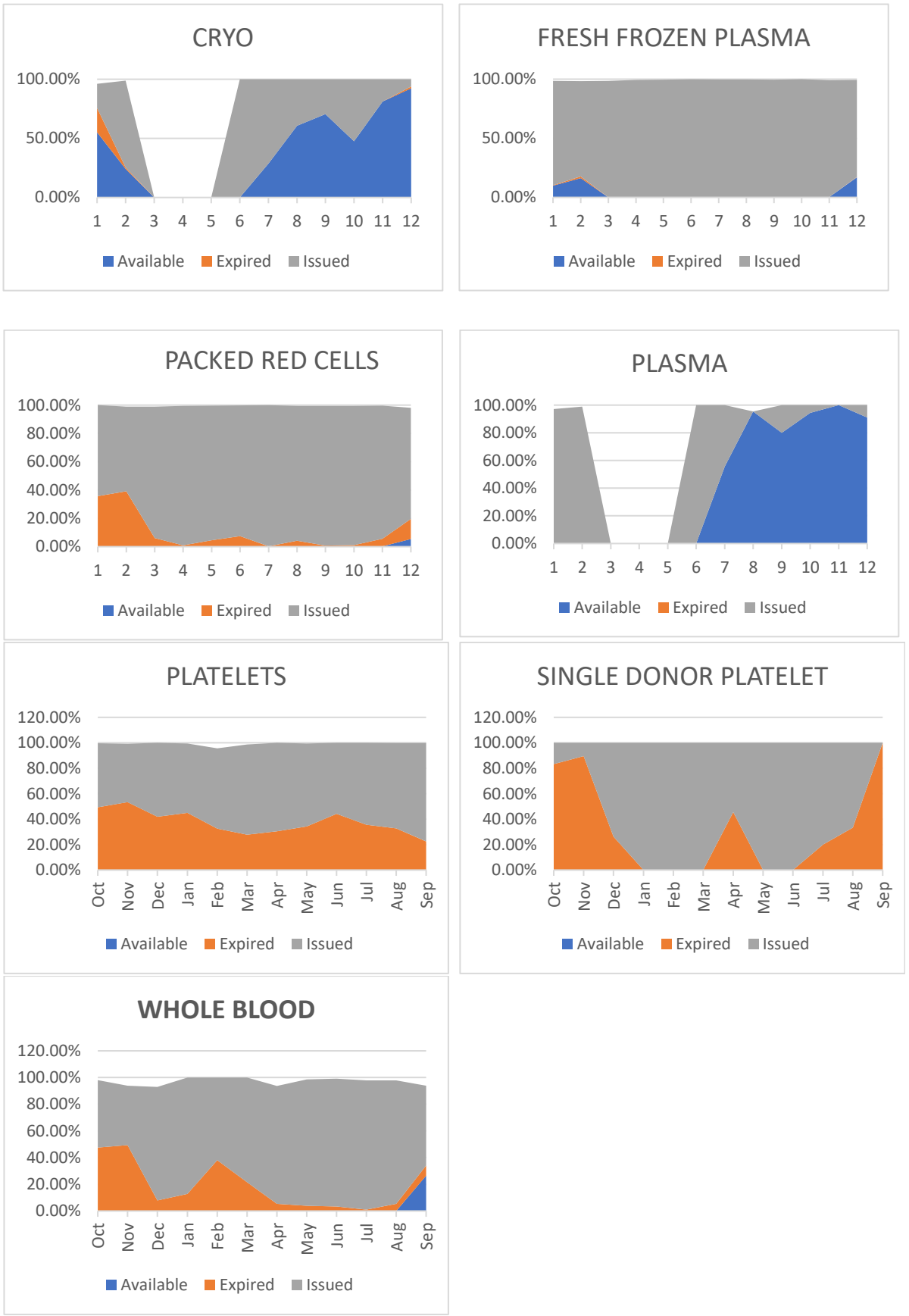

Figure: Product utilization, Expiry and availability distributions wrt its supply over months

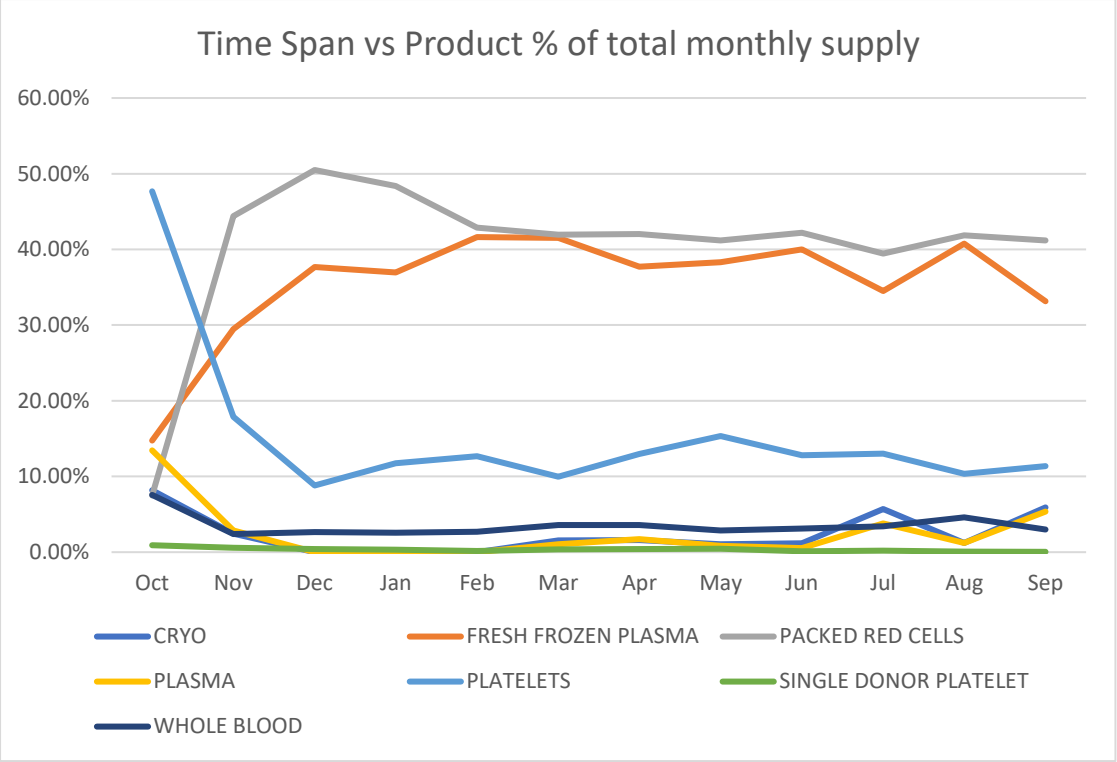

Figure: Product Supply Visualization

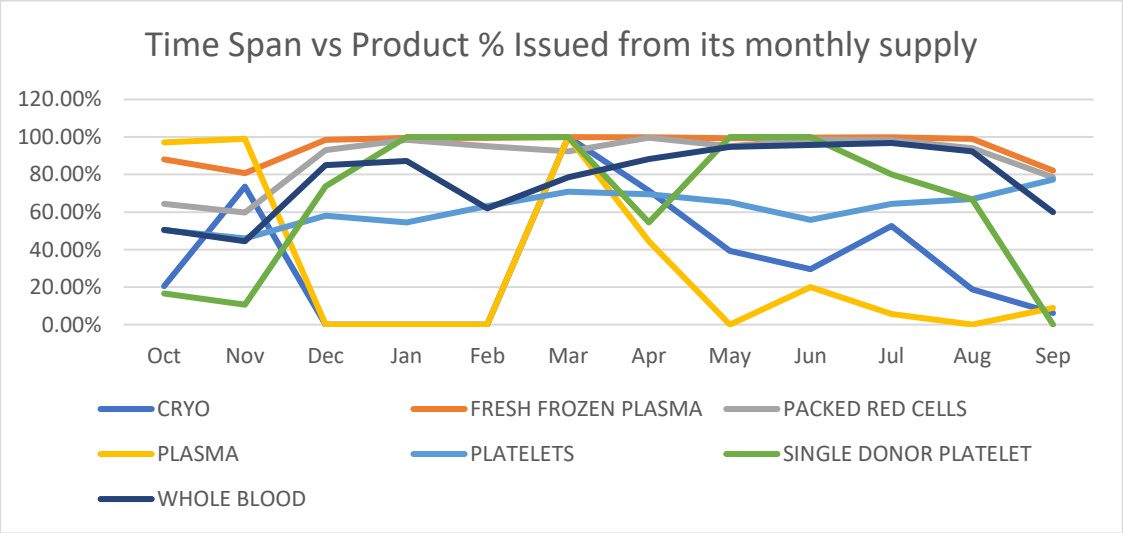

Figure: Product Demand Visualization

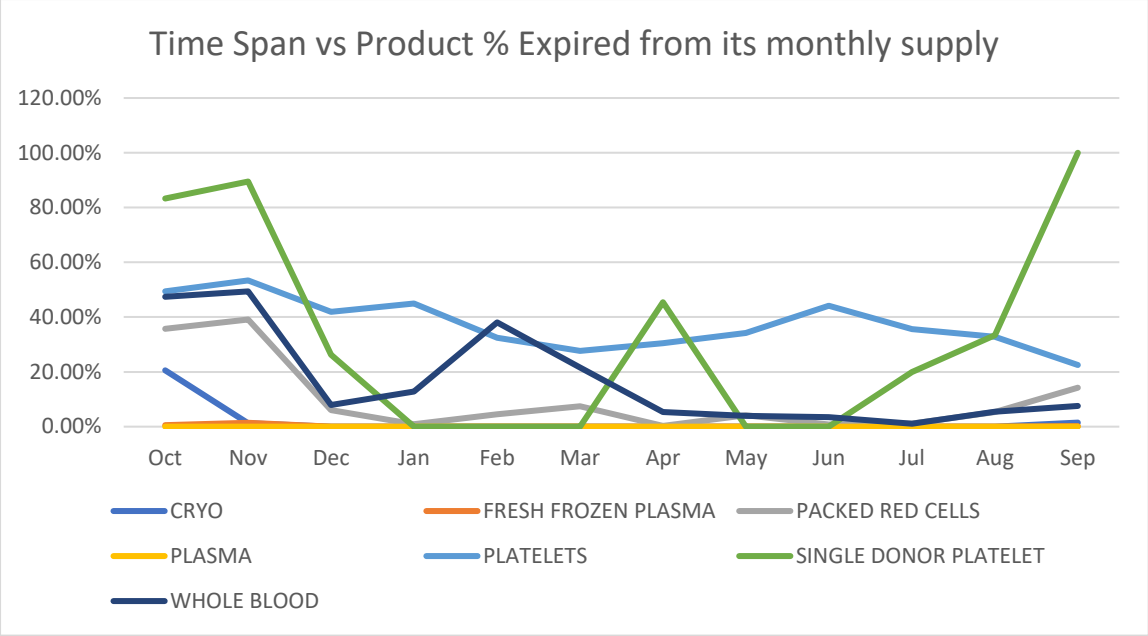

Figure: Product Wastage Visualization

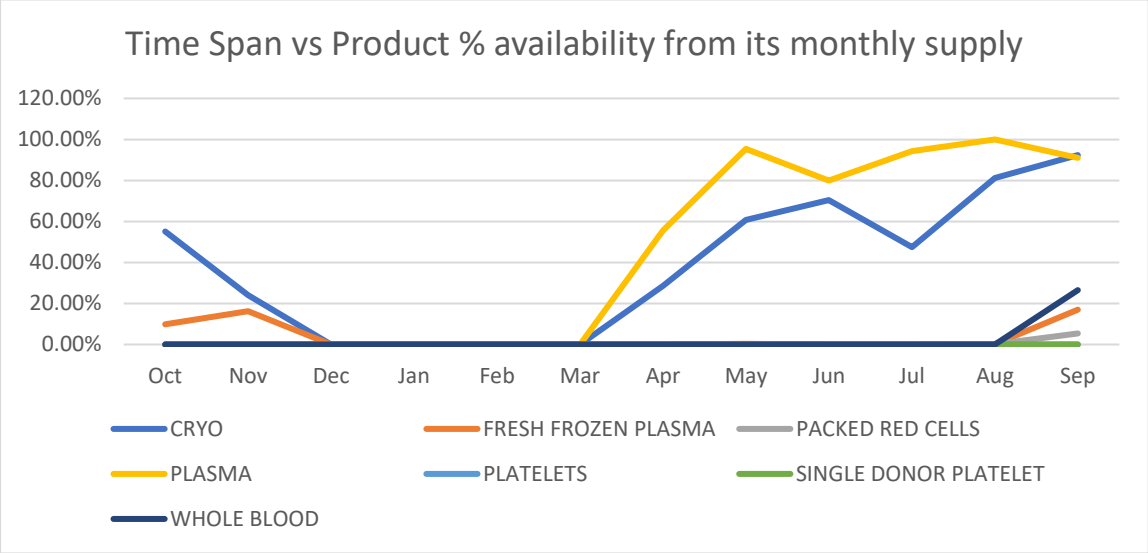

Figure: Product Extra Availability Visualization

These charts give some interesting insights into how each product is utilized every month and also suggests their wastages due to expiry.

2.2. Is there any abnormality or pattern in your visualizations or statistics? Is your observation statistically significant? Perform at least one proper hypothesis test and explain the rationale behind the test.

Through these visualizations, we could understand some important patterns in the bags supply, demand and wastages over different months, as listed below:

1. October '17 and November '17 months have highest % of supplied bags expired (wasted).
2. Sept '18 month has a huge gap between supply and demand of bags
3. About 38 % of the supplied "Platelets" and "Single Donor Platelets" gets expired, which is much higher rate than other products
4. Almost over the entire year, "Fresh Frozen plasma" and "packed red cells" accounts for the highest number of bags supply received at VHS. Also, these 2 products are consistently issued at higher rates and have very less wastages or surplus availability.
5. "platelets" supply reduces drastically from oct '17 to sept '18
6. "plasma" and "cryo" have huge fluctuations in their supply and demand over the year. This also explains why they are available extra in high numbers for many months. So, to avoid associated storage costs, VHS can develop better strategy around its supply.

So, we can now check our above claim no. 3 with statistical hypothesis testing:

Null Hypothesis: H0: Expiry or wastage rate is same for all products

Alternative Hypothesis: H1: Expiry rate for at least one product is different from other products

Expiry rates	CRYO	FRESH FROZEN PLASMA	PACKED RED CELLS	PLASMA	PLATELETS	SINGLE DONOR PLATELET	WHOLE BLOOD
Oct	20.56%	0.52%	35.71%	0.00%	49.36%	83.33%	47.47%
Nov	1.20%	1.40%	39.12%	0.00%	53.39%	89.47%	49.38%
Dec	0.00%	0.00%	6.01%	0.00%	41.98%	26.32%	7.87%
Jan	0.00%	0.00%	0.93%	0.00%	45.01%	0.00%	12.79%
Feb	0.00%	0.00%	4.50%	0.00%	32.46%	0.00%	38.05%
Mar	0.00%	0.00%	7.44%	0.00%	27.70%	0.00%	21.54%
Apr	0.00%	0.00%	0.27%	0.00%	30.50%	45.45%	5.32%
May	0.00%	0.00%	4.22%	0.00%	34.24%	0.00%	3.95%
Jun	0.00%	0.00%	0.77%	0.00%	44.09%	0.00%	3.45%
Jul	0.00%	0.00%	1.10%	0.00%	35.56%	20.00%	1.06%
Aug	0.00%	0.00%	5.30%	0.00%	32.77%	33.33%	5.43%
Sep	1.53%	0.20%	14.18%	0.00%	22.51%	100.00%	7.58%

By applying one-way Anova test to check if these column groups values differ or not, we get following results:

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
CRYO	12	0.23292284	0.01941	0.00346651
FRESH FROZEN PLASI	12	0.02127126	0.001773	1.732E-05
PACKED RED CELLS	12	1.19548396	0.099624	0.01796765
PLASMA	12	0	0	0
PLATELETS	12	4.49564355	0.374637	0.00859103
SINGLE DONOR PLAT	12	3.97910686	0.331592	0.14568475
WHOLE BLOOD	12	2.03902899	0.169919	0.03190161

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.769941528	6	0.29499	9.9453019	4.14754E-08	2.218817
Within Groups	2.283917555	77	0.029661			
Total	4.053859084	83				

As p-value is less than 0.05, we need to reject our null hypothesis in the favour of alternative hypothesis. So, our claim that "Platelets" and "Single Donor Platelets" got expired and wasted at much higher rate than other products is statistically consistent.

2.3. Based on your findings from the previous parts, do you have any suggestions for VHS to improve its process?

Based on our findings listed in section 2.2, I have following suggestions for VHS to improve its processes, and strategies:

VHS receives blood donation and processes into different blood components before storing and issuing for patient uses. So, we have observed that certain blood products do not have that much requirements compared to others, and demand also varies over the months. Also, supply from donors from donation camps also varies over the year. Therefore following steps can be taken based on current gaps in supply, and demand to have just-in-time (JIT) model to avoid costs related to wastages due to expiry and storage costs over extended periods. Particularly following actions shall be taken after detailed investigation on the underlying reasons:

- Bags Wastage due to expiry in October and November months were the highest. So, planning supply and donation drives to match with demand in those months is highly recommended.
- September month has a huge gap between supply and demand of bags, so better planning of supply is required for that month as well.
- About half of the "Platelets" and "Single Donor Platelets" collected from donation and processed got expired due to less demands, so reducing their supply to match with demand is highly encouraged.
- "Fresh Frozen plasma" and "packed red cells" products have been doing well in terms of match between supply and demand quantities and results in very less percentage wastages.
- "platelets" supply volatility to be reduced as some months did not have any platelets collected and available for issue.
- "plasma" and "cryo" have huge fluctuations in their supply and demand over the year. So, accurate demand forecasting would help VHS in reducing its storage and extra availability related costs.

References:

[1] H A Mehrizi, eBook: MSCI 719 Winter 2023 Cases Multiple (ID: 9723713) Accessed: Jan. 22, 2023. [Online].

Available: https://www.campusebookstore.com/integration/AccessCodes/default.aspx?permalinkId=ee044bf2-fe82-4db0-ad22-088e81954eef&frame=YES&t=permalink&sid=4u2faw45zyslbp45bbqlpc55