Conducting Hypothesis Testing on supply chain management dataset

```
In [1]:
          import numpy as np
          import pandas as pd
In [2]:
          import warnings
          warnings.filterwarnings('ignore')
In [3]:
          df=pd.read_csv("C:\\Users\\vamsi\\OneDrive\\Desktop\\Vamshi Data\\Supplychain train
In [4]:
          df.head(10)
Out[4]:
            Ware_house_ID WH_Manager_ID Location_type WH_capacity_size
                                                                          zone WH_regional_zone
         0
               WH_100000
                                 EID_50000
                                                  Urban
                                                                   Small
                                                                          West
                                                                                          Zone 6
         1
               WH_100001
                                 EID_50001
                                                  Rural
                                                                         North
                                                                                          Zone 5
                                                                   Large
         2
               WH_100002
                                 EID_50002
                                                                                          Zone 2
                                                  Rural
                                                                    Mid
                                                                         South
         3
               WH_100003
                                 EID_50003
                                                  Rural
                                                                    Mid
                                                                         North
                                                                                          Zone 3
               WH 100004
                                 EID_50004
         4
                                                  Rural
                                                                   Large
                                                                         North
                                                                                          Zone 5
         5
               WH 100005
                                 EID 50005
                                                  Rural
                                                                   Small
                                                                          West
                                                                                          Zone 1
         6
               WH 100006
                                 EID_50006
                                                                                          Zone 6
                                                  Rural
                                                                   Large
                                                                          West
         7
               WH 100007
                                 EID_50007
                                                  Rural
                                                                         North
                                                                                          Zone 5
                                                                   Large
         8
               WH 100008
                                 EID 50008
                                                  Rural
                                                                   Small
                                                                         South
                                                                                          Zone 6
               WH 100009
                                 EID 50009
                                                  Rural
                                                                                          Zone 6
                                                                   Small South
        10 rows × 24 columns
In [5]:
         df.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 22150 entries, 0 to 22149
         Data columns (total 24 columns):
          #
              Column
                                              Non-Null Count Dtype
                                               -----
          0
              Ware house ID
                                              22150 non-null object
              WH_Manager_ID
                                              22150 non-null object
          1
          2
                                              22150 non-null object
              Location_type
          3
              WH_capacity_size
                                              22150 non-null
                                                               object
          4
                                              22150 non-null
                                                               object
              zone
          5
                                                               object
              WH_regional_zone
                                              22150 non-null
              num refill_req_13m
          6
                                              22150 non-null
                                                               int64
          7
              transport_issue_l1y
                                              22150 non-null
                                                               int64
              Competitor_in_mkt
          8
                                              22150 non-null
                                                               int64
              retail shop num
                                              22150 non-null
                                                               int64
```

```
10 wh_owner_type
                                22150 non-null object
11 distributor_num
                                22150 non-null int64
12 flood_impacted
                                22150 non-null int64
13 flood_proof
                                22150 non-null int64
14 electric_supply
                                22150 non-null int64
                                22150 non-null int64
15 dist_from_hub
16 workers_num
                                21273 non-null float64
17 wh_est_year
                                11605 non-null float64
18 storage_issue_reported_l3m 22150 non-null int64
19 temp_reg_mach
                               22150 non-null int64
20 approved_wh_govt_certificate 21345 non-null object
21 wh_breakdown_13m
                                22150 non-null int64
22 govt_check_13m
                                22150 non-null int64
23 product_wg_ton
                                22150 non-null int64
dtypes: float64(2), int64(14), object(8)
memory usage: 4.1+ MB
```

In [6]:

df.dtypes

Out[6]:

Ware_house_ID	object
WH_Manager_ID	object
Location_type	object
WH_capacity_size	object
zone	object
WH_regional_zone	object
num_refill_req_13m	int64
transport_issue_l1y	int64
Competitor_in_mkt	int64
retail_shop_num	int64
wh_owner_type	object
distributor_num	int64
flood_impacted	int64
flood_proof	int64
electric_supply	int64
dist_from_hub	int64
workers_num	float64
wh_est_year	float64
storage_issue_reported_13m	int64
temp_reg_mach	int64
<pre>approved_wh_govt_certificate</pre>	object
wh_breakdown_13m	int64
<pre>govt_check_13m</pre>	int64
product_wg_ton	int64
dtype: object	

Descrpitive Stats

In [7]: df.describe().T

Out[7]:		count	mean	std	min	25%	50%	75%
	num_refill_req_l3m	22150.0	4.097020	2.606289	0.0	2.00	4.0	6.0
	transport_issue_l1y	22150.0	0.777201	1.201747	0.0	0.00	0.0	1.0
	Competitor_in_mkt	22150.0	3.103928	1.142886	0.0	2.00	3.0	4.0
	retail_shop_num	22150.0	4983.115711	1050.634225	1821.0	4309.25	4859.0	5499.0
	distributor_num	22150.0	42.386998	16.057730	15.0	29.00	42.0	56.0
	flood_impacted	22150.0	0.098691	0.298253	0.0	0.00	0.0	0.0

		count	mean	std	min	25%	50%	75%
	flood_proof	22150.0	0.054492	0.226991	0.0	0.00	0.0	0.0
	electric_supply	22150.0	0.656072	0.475028	0.0	0.00	1.0	1.0
	dist_from_hub	22150.0	163.613725	62.660709	55.0	109.00	164.0	218.0
	workers_num	21273.0	28.936398	7.843431	10.0	24.00	28.0	33.0
	wh_est_year	11605.0	2009.401206	7.527223	1996.0	2003.00	2009.0	2016.0
	storage_issue_reported_I3m	22150.0	17.116659	9.174193	0.0	10.00	18.0	24.0
	temp_reg_mach	22150.0	0.304199	0.460078	0.0	0.00	0.0	1.0
	wh_breakdown_l3m	22150.0	3.487765	1.691661	0.0	2.00	3.0	5.0
	govt_check_l3m	22150.0	18.767765	8.644778	1.0	11.00	20.0	26.0
	product_wg_ton	22150.0	22086.780813	11626.192340	2065.0	12151.00	22099.0	30102.0
	1							•
In [8]:	df.shape							
Out[8]:	(22150, 24)							
In [9]:	#checking missing valu	es:						
	df.isnull().sum()							
Out[9]:	Ware_house_ID WH_Manager_ID Location_type WH_capacity_size zone WH_regional_zone num_refill_req_l3m transport_issue_l1y Competitor_in_mkt retail_shop_num wh_owner_type distributor_num flood_impacted flood_proof electric_supply dist_from_hub workers_num wh_est_year storage_issue_reported_ temp_reg_mach approved_wh_govt_certif wh_breakdown_l3m govt_check_l3m product_wg_ton dtype: int64		0 0 0 0 0 0 0 0 0 0 0 0 877 10545 0 0 805 0					
In [10]:	<pre>#Boolean Output: df.isnull().any()</pre>							
Out[10]:	Ware_house_ID WH_Manager_ID		False False					

False Location_type False WH_capacity_size zone False WH_regional_zone False num refill req 13m False transport_issue_l1y False Competitor_in_mkt False retail_shop_num False wh_owner_type False distributor_num False flood_impacted False flood_proof False electric_supply False dist_from_hub False workers_num True wh_est_year True storage_issue_reported_13m False temp_reg_mach False approved_wh_govt_certificate True wh_breakdown_13m False govt_check_13m False product_wg_ton False dtype: bool

In [11]:

df.isna().apply(pd.value_counts).T

Out[11]:

	False	True
Ware_house_ID	22150.0	NaN
WH_Manager_ID	22150.0	NaN
Location_type	22150.0	NaN
WH_capacity_size	22150.0	NaN
zone	22150.0	NaN
WH_regional_zone	22150.0	NaN
num_refill_req_l3m	22150.0	NaN
transport_issue_l1y	22150.0	NaN
Competitor_in_mkt	22150.0	NaN
retail_shop_num	22150.0	NaN
wh_owner_type	22150.0	NaN
distributor_num	22150.0	NaN
flood_impacted	22150.0	NaN
flood_proof	22150.0	NaN
electric_supply	22150.0	NaN
dist_from_hub	22150.0	NaN
workers_num	21273.0	877.0
wh_est_year	11605.0	10545.0
storage_issue_reported_I3m	22150.0	NaN
temp_reg_mach	22150.0	NaN

	False	True
approved_wh_govt_certificate	21345.0	805.0
wh_breakdown_I3m	22150.0	NaN
govt_check_l3m	22150.0	NaN
product_wg_ton	22150.0	NaN

```
In [12]: #percentage of missing values
    a=df.isna().sum()
    perc=(a/len(df))*100
```

```
In [13]: perc
```

Ware house ID 0.000000 Out[13]: WH_Manager_ID 0.000000 Location type 0.000000 WH_capacity_size 0.000000 zone 0.000000 WH_regional_zone 0.000000 num_refill_req_13m 0.000000 transport_issue_l1y 0.000000 Competitor_in_mkt 0.000000 retail_shop_num 0.000000 wh_owner_type 0.000000 distributor num 0.000000 flood_impacted 0.000000 flood_proof 0.000000 electric_supply 0.000000 dist_from_hub 0.000000 workers_num 3.959368

 approved_wh_govt_certificate
 3.634312

 wh_breakdown_l3m
 0.000000

 govt_check_l3m
 0.000000

 product_wg_ton
 0.000000

 dtype: float64

storage_issue_reported_13m

import matplotlib.pyplot as plt
import seaborn as sns

In [15]: corr=df.corr() corr

wh_est_year

temp_reg_mach

Out[15]: num_refill_req_l3m transport_issue_l1y Competitor_in_mkt retail_shop_nu num_refill_req_l3m 1.000000 0.020336 0.008104 -0.0017 transport_issue_l1y 0.020336 1.000000 -0.008288 0.0013 Competitor_in_mkt 0.008104 -0.008288 1.000000 -0.1548 retail_shop_num -0.001776 0.001391 -0.154848 1.0000 distributor_num 0.005063 0.005509 0.000886 -0.0043

47.607223

0.000000

0.000000

			ı	num_r	efill_r	eq_l3n	n tr	anspo	ort_iss	ue_l1	у С	ompet	itor_i	n_mk	t re	tail_shop_nu
	flood	impac	ted		-0.	.01458	3		-0.	00667	1		0.0	08982	2	-0.0053
	flo	od_pr	oof		-0.	.00481	6		0.	00313	1		-0.0	04428	8	0.0045
	elect	ric_sup	ply		-0.	.00721	7		-0.	00737	8		-0.0	000688	8	-0.0068
	dist_	from_l	nub		-0.	.00061	1		0.	01386	4		0.0	10358	8	0.0019
	wor	kers_n	um		-0.	.01375	1		-0.	00962	.5		-0.0	04387	7	-0.0072
	wl	n_est_y	ear		0.	.01262	8		-0.	01383	7		-0.0)1210 ⁻	1	0.0076
storage_is	sue repo	orted I	3m		-0.	.00771	9		-0.	14591	4		0.0	07654	4	-0.0070
3 -		reg_m				.26353				01275)11397		-0.0004
w	h_break	_				.000704				01179				1254		-0.0056
	govt_c					.00647(00605)43094		0.0435
	produc					.00047				17574				07282		-0.0067
	produ	ut_wg_	COII		0.	.000071	0		0.	17574	.0		0.0	707202	_	0.0007
◀																•
plt.fig plt.tit	ure(fig le(' <mark>Co</mark> r				x',fo	ontsi	ze=2	.0)								
	le(' <mark>Cor</mark> tmap(co	relat	ion_N	Matri:		nnot=	True	, vma		∂,∨mi	.n=1.	0,fmi	t='g')		
plt.tit sns.hea	le(' <mark>Cor</mark> tmap(co	relat	ion_N	Matri:		nnot=	True			0,∨mi	.n=1.	0,fm1	.000704451	.0.00647035	0.0006757	1100
plt.tit sns.hea plt.sho	le('Cor tmap(co w()	relat	cion_N nap='F	Matri:	n', ar	nnot='	True	Vma on_Matri	x				0000704451		0.0006757	-1100
plt.tit sns.hea plt.sho rum_refil_req_Bm. transport_issue_Ily. Competitor_in_mat.	le('Cor tmap(cow()	orr, cm	cion_N nap='F	Aatri RdYlG 00056316 00055946	n', ar	nnot='	True	Vma on_Matri	x				0000704451 00117954 0012544	0.00647035 0.00605783 -0.0430942	-0.175744 0.007282	4
plt.tit sns.hea plt.sho rum.refil.req.l3m. bransport.jssue_l1y. Campetitor_in_mat- retail_shop_num.	le('Cor tmap(cc w()	orr, cm	o.00177582	Matri:	o 0145835 -0 00667122 -0 00698208	nnot='	True	Vma on_Matri	x		-0.00771853 -0.145914	0.263539	0.000704451 0.0117954 0.012544 0.00569419		-0.175744 0.007282 -0.006785	4
plt.tit sns.hea plt.sho rum_refil_req_lim. transport_issue_lly. Competitor_in_mit. retail_shop_rum.	le('Cortmap(cow()) 1 00203364 1 000810396 -0.00828 0.00177582 0.001396	orr, cm	cion_Nap='F	Aatri RdYlG 00056316 00055946	n', ar	nnot='	True	Vma on_Matri	x	0.0126284 -0.0138365 -0.012101 -0.0076268 -0.0153747	-0.00771853 -0.145914	0.263539 0.0127567 0.0113972 0.000463717 0.00404815	0.000704451 0.0117954 0.012544 -0.00569419 0.00188458	0.00647035 0.00605783 -0.0430942	-0.175744 0.007282 -0.006785 0.005142	-1075
plt.tit sns.hea plt.sho rum_refil_req_l3m transport_issue_l1y Competitor_in_mixt retail_shop_rum distributor_rum-	le('Cortmap(cow()) 1 00203364 1 000203364 1 000810396 -0.008285 0.001396	onn, cm 4 00010396 -0.00028765 1 0.0008597; 22 0.00098208	a 00177582 0 000139083 0 154848 1 0 000439612 0 00053205	Adri: AdylG 00050316 00050946 0000085971 000439612	0.0145835 -0.00667122 0.00898208 -0.0089038	000401579 0000412756 4000402756 4	True prrelatio 0.00721686 0.00737766 0.000688437 0.000681161	on_Matri. 0.006611218 0.0138636 0.0103577 0.00198277	X -0.0137506 -0.00962509 -0.00439717 -0.00728086 -0.0143709	0.0126284 -0.0138365 -0.012101 0.0076268 -0.0153747	-0.00771853 -0.145914 0.00765372 -0.00704518 0.0033139	0.263539 0.0127567 0.0113972 0.000463717 0.00404815	0.000704451 0.0117954 0.012544 0.00569419 0.00188458 -2.28014e.05	.0.00647035 0.00605783 -0.0430942 0.0435573 -0.00599627 -0.00117909	-0.175744 0.007282 -0.006785 0.005142	-1075 -1050 -1025
plt.tit sns.hea plt.sho num_refil_req_llm. transport_issue_lly. Competior_in_mixt. retail_shop_num. dstributor_num. flood_impacted. flood_proof.	le('Cortmap(cow()) 1 00203364 1 000203364 1 000610396 0001396 0005506316 0005506	0 00010396 4 0 00010396 54 0 00010396 55 1 83 0 154048 46 0 0008597 22 0 00095206	anap='F	0.00506316 0.00550946 0.00439612 1 0.00609553	0 0145835 0 00667122 0 0069208 0 0069038	CCC 0.00481579 0.00031312 0.000450235 0.00069553 0.0066522	True orrelatio 0.00731766 0.000688437 0.00068161 0.00076266 0.167565	on_Matri, 0.00611218 0.013836 0.0103577 0.00198277 -0.00753793	X	0.0126284 -0.0138365 -0.012101 -0.0076268 -0.0153747 -0.00186629	.0.00771853 .0.145914 .0.00765372 .0.00704518 .0.0033139 .0.00622091	0.263539 0.0127567 0.0113972 0.00404815 -0.0118378	0.000704451 0.0117954 0.012544 0.00569419 0.00188458 -2.28014e.05	.0.00647035 0.00605783 -0.0430942 0.0435573 -0.00599627 -0.00117909	-0.175744 0.007282 -0.006785 0.0051429	-1075 -1050 -1025
plt.tit sns.hea plt.sho rum.refil.req.l3m. bransport.jssue_l1y. Campetitor_in_mat. retail_shop_num. distributor_num. flood_impacted. flood_proof. dectric_supply. dist_from_hub	le('Cortmap(ccum)) 1 00203364 1 00203364 1 000510396 0.00825 0.00177582 0.001396 0.005505 0.0145835 0.00667 0.00451579 0.00313 0.007377 0.000611218 0.01386	0.00810396 4.000810396 4.00082876 55 1 83 4.0154848 4.000885977 6.000442756 6.0006843 6.0006843	a 00177582 0 00139083 0 154848 1 0 00459235 7 0 00459235 7 0 00681161 0 00190277	Adrix AdYlG 0.0056316 0.00550446 0.00439612 1 0.0060953 0.0060953 0.0076266 0.00753793	0 0145835 0 00667122 0 00698208 0 0053205 0 0088038 1 0 106622 0 167565	000481579 00031312 000442756 000442756 000442756 000442756 000442756 000450235 000442756 000450235 00045025 000450000000000	Orrelation	on_Matrii 0 000611218 0 0138636 0 0103577 0 00198277 -0 00753793 0 00562017 -0 00318621 -0 00213817	X	0.0126284 -0.0138365 -0.012101 0.0076268 -0.0153747 0.001212171 -0.00186629 -0.00676566 0.009228	-0.00771853 -0.145914 -0.00765372 -0.00704518 -0.0033139 -0.00622091 -0.00506906 -0.00623062	0.263539 0.0127567 0.0113972 -0.000463717 0.00404815 -0.0118378 0.00425067 -0.00632449	0.000704451 0.0117954 0.012544 -0.00569419 0.00188458 -2.28014e.05 -0.00129965	-0.00647035 0.00605783 -0.0430942 0.0435573 -0.00399627 -0.00117909 -0.00717149 -0.00725813	-0.175744 0.007287 -0.006785 0.0051429 -0.0050555 -0.0035668 -0.0007254 -0.0060400	-1075 -1050 -1025 9
plt.tit sns.hea plt.sho num.refil.req.lbm. transport.issue_lly. Competitor_in_mixt. retail_shop_num. distributor_num. flood_impacted. flood_impacted. dectric_supply. dst_from_hub. workers_num.	le('Cortmap(ccow()) 1 00203364 1 000203364 1 000810396 0.008285 000177582 0.001396 000506316 0.005506 0004015279 0.00313 0006077 0006011218 0.01386 0.0137506 0.00962	0.00828755 0.00828755 0.00828755 1.00885977 0.00885977 0.00885977 0.0088597	a 00177582 a 00177582 a 00139083 a 0.154848 1 a 0.0039057 a 0.00439612 b 0.00439612 c 0.00490235 c 0.00450235 c 0.00450235	000506316 000550946 0000885971 0000038 0000038 0000038 0000038	0.0145835 -0.00667122 0.00698208 -0.0053205 0.0080038 1 0.106622 0.167565 0.00562017	000481579 00031312 000442756 4 000450235 0106622 1 0115616 000318621 000410307	Optical and Control of	on_Matri, 0.00611218 0.0138636 0.0103577 0.00198277 0.00753793 0.00562017 0.00318621 1 0.00213817	X	0.0126284 -0.0138365 -0.012101 -0.0076268 -0.0153747 -0.00212171 -0.00186629	.0.00771853 .0.145914 .0.00765372 .0.00704518 .0.0052091 .0.0052091 .0.00521449 .0.00653062 .0.00878818	0.263539 0.0127567 0.0113972 0.00463717 0.0044815 -0.0118378 0.00425067 -0.00632449 0.00017497	0.000704451 0.0117954 0.012544 0.00569419 0.00188458 2.28014e-05 0.00796382 0.00129965 0.00129965	.0.00647035 0.00605783 .0.0430942 0.0435573 .0.00599627 .0.00117909 .0.00717149 .0.00725813 0.00276858	-0.17574i 0.007282 -0.006785 0.005142: -0.0050552 -0.0035668 -0.007254 -0.0060400	-1075 -1050 -1025 9
plt.tit sns.hea plt.sho rum.refill.req.lim. transport_issue_liy. Competitor_in_mixt. retail_shop_rum. distributor_rum. flood_impacted. flood_impacted. doct_ircs_supply dst_from_rub workers_rum wh_est_year	le('Cortmap(ccum)) 1 00203364 1 00203364 1 000510396 0.00825 0.00177582 0.001396 0.005505 0.0145835 0.00667 0.00451579 0.00313 0.007377 0.000611218 0.01386	0.00010396 0.00028765 1.000888973 0.154848 0.00088971 0.00088973 0.00088973 0.00088973 0.00088973	anap='F anap='	0.00506316 0.00506316 0.00506346 0.000885971 0.00439612 1 0.0060038 0.00076266 0.00753793 0.0143709	0 0145835 0 00667122 0 00698208 0 0053205 0 0088038 1 0 106622 0 167565	000481579 00031312 000442756 4 000450235 0106622 1 0115616 000318621 000410307	Occident	on_Matri	X	0.0126284 -0.0138365 -0.012101 -0.0076268 -0.0153747 -0.00186629 -0.00676566 -0.009228 -0.0066781	-0.00771853 -0.145914 -0.00765372 -0.00704518 -0.0033139 -0.00622091 -0.00506906 -0.00623062	0.263539 0.0127567 0.0113972 -0.000463717 0.00404815 -0.0118378 0.00425067 -0.00632449	0.000704451 0.0117954 0.012544 -0.00569419 0.00188458 -2.28014e.05 -0.00129965	-0.00647035 0.00605783 -0.0430942 0.0435573 -0.00399627 -0.00117909 -0.00717149 -0.00725813	-0.175744 0.007287 -0.006785 0.0051429 -0.0050555 -0.0035668 -0.0007254 -0.0060400	-1075 -1050 -1025 -1000
plt.tit sns.hea plt.sho rum.refill.req.lim. transport_issue_liy. Competitor_in_mixt. retail_shop_rum. distributor_rum. flood_impacted. flood_impacted. doct_ircs_supply dst_from_rub workers_rum wh_est_year	le('Cortmap(cou) 1 002033 00203364 1 000610396 000628 00017582 0001396 000566316 0005566 00145835 000667 00145835 000667 0006611218 0001386 00137506 000962	00010396 4 00010396 4 0008897 5 1 3 0154948 6 0008897 2 0.00942756 6 0.00068843 6 0.0005377	a. 0.00177592 a. 0.00177592 b. 0.00199083 a. 0.0154848 b. 1 a. 0.00439612 c. 0.00450235 c. 0.0045025 c. 0.0045025 c. 0.0045025 c	Adrix RdYlG 0.00506316 0.00550946 0.00685971 0.00439612 1 0.00609553 0.00609553 0.00676266 0.00753793 0.0143709 0.0153747	0.0145835 -0.00667122 0.00698208 -0.0053205 0.008038 -1 0.106622 0.167565 0.00952017 0.166148	000481579 00031312 000442756 4000450235 000699533 000699533 00069950 00069950 000690000000000	Occident	on_Matri	X	0.0126284 -0.0138365 -0.012101 -0.0076268 -0.0153747 -0.00186629 -0.00676566 -0.009228 -0.0066781	.0.00771853 -0.145914 -0.00765372 -0.00704518 -0.0033139 -0.00522091 -0.00506906 -0.00653062 -0.00678818 -0.858409	0.263539 0.0127567 0.0113972 40.000463717 0.00404915 -0.0116378 0.00425067 -0.00632449 0.00017497 0.00017497	0.000704451 0.0117954 0.012544 0.00569419 0.00188458 2.28014e-05 0.00796382 0.00129963 0.000129963	.0.00647035 0.00605783 .0.0430942 0.0435573 .0.00599627 .0.00117909 .0.00717149 .0.00725813 .0.00276858 .0.00491306 .0.00836377	-0.175744 0.007282 -0.006785 0.005142: -0.0050552 -0.0035666 -0.007254 -0.006400 -0.628743	-1075 -1050 -1025 -1000
plt.tit sns.hea plt.sho num_refil_req_l3m. transport_issue_lly. Competitor_in_mixt. retail_shop_num. distributor_num. flood_impacted. flood_groof. dectric_supply. dst_from_nub. west_year. dorage_issue_reported_l3m. temp_reg_mach.	le('Cortmap(ccow()) 1 00203364 1 00203364 1 000810396 0.00828 0.00177582 0.001395 0.00506316 0.005505 0.00491579 0.00313 0.00721686 0.007377 0.000611218 0.01386 0.0137506 0.00962 0.0126264 0.01385	0.00810396 0.00810396 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 1 0.00828765 0.0	a.00177592 a.00177592 a.00139083 a.0154848 1 a.0053205 a.000439612 a.0053205 a.000450235 7 a.000580161 a.00198277 a.00752886 a.00752886 a.000463717	Aatrix RdYlG 0.0050946 0.00559946 0.00439612 1 0.0060953 0.00609553 0.006753793 0.0143709 0.0153747 0.0033139	0.0145835 0.00667122 0.0069208 0.0053205 0.0060038 1 0.106622 0.167565 0.00562017 0.166148 0.005212171 0.00622091 0.00622091	000415027 4 000415067 4 000425067 4	True 0.00721696 0.00731766 0.007031766 0.00068161 0.00068161 1 0.0013817 0.0013817 0.0017556	0013837 00052017 000318217 00038277 000198277 00038277 000318217 000318217 1 000323817	X	0.0126284 0.0138365 -0.012101 0.0076268 -0.0153747 0.00212171 -0.00186629 -0.00676566 0.009228 0.00666781	.0 00771853 -0.145914 -0.00765372 -0.00704518 -0.0033139 -0.00622091 -0.000621449 -0.00653062 -0.00878818 -0.858409	0263539 00127567 00113972 0.004063717 0.00404015 -0.0118378 0.00425067 -0.00632449 0.00017497 0.000716717 -0.0463952	0.000704451 0.0117954 0.012544 0.00569419 0.00188458 2.28014e-05 0.00796382 0.0124282 0.00129965 -0.0022511 0.398383	.0.0047035 0.00695783 .0.0430942 0.0435573 .0.0017909 .0.00717149 .0.00725813 0.00276658 .0.00491306 0.00836377 .0.010313	-0.175746 0.007287 -0.006785 0.005142; -0.0050552 -0.0035666 -0.0060400 -0.0060400 -0.0828743	-1075 -1050 -1025 -1000
plt.tit sns.hea plt.sho num.refil req.lim. transport.issue_lly. Competitor_in_mixt. retail_shop_num. dstributor_num. flood_impacted. flood_impacted. flood_impacted. flood_impacted. strib_imparted. strib_imparted.imparte	le('Cortmap(ccortmap(ccortmap)) 1 00203364 1 000203364 1 0008283 0001396 0005566 0005566316 0005566 000481579 000313 000721686 000737 000481579 000313 000721686 000737 000667 000677 000677 000677 000737 000677 00073	0.00810396 0.00810396 0.00828765 1.00810396 0.0088977 22.00088937 22.00088937 23.00088937 24.0003877 25.0003877 26.00068843 26.00068843 27.00068843 28.00068843 29.00068843 20.00068843	a.00177592 a.00177592 a.00139083 a.0154848 1 a.0053205 a.000439612 a.0053205 a.000450235 7 a.000580161 a.00198277 a.00752886 a.00752886 a.000463717	000506316 000550946 000550946 0000885971 000439612 1 00060038 000609553 000609553 00076266 000753793 00153747 00033139 00044815 00044815	0.0145835 -0.00667122 -0.00697122 -0.0053205 -0.0053205 -0.0069238 -1 -0.106622 -0.167565 -0.00552017 -0.166148 -0.00522017 -0.00622091 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.006220 -0.00620 -0.006220	000481579	True Dorrelatio 0.00731766 0.00731766 0.00688437 0.00688437 0.00688437 0.00681161 1 0.00731866 1 0.00731866 0.00731866 0.00731867 0.00821449 0.00632449 0.00672583	0.00562017 0.00213817 0.00318621 0.00318621 0.00318621 0.00318621 0.00318621 0.00318621 0.00318621 0.00318621 0.00318621	X	0.0126284 -0.0138365 -0.012101 -0.0076268 -0.0153747 -0.00212171 -0.00186629 -0.00676566 -0.009228 -0.00686781 -1 -0.858409 -0.0463952 -0.398383	.0.00771853 -0.145914 0.00765372 -0.00704518 0.0033139 -0.00622091 -0.00506906 -0.00678818 -0.858409 -1 -0.00539011	0.263539 0.0127567 0.0113972 -0.00463717 0.00404815 -0.0118378 0.00425067 -0.00632449 0.00017497 0.00632449 0.00632449	0.000704451 0.0117954 0.012544 0.00569419 0.00188458 2.28014e-05 0.00796382 0.0124282 0.00129965 -0.0022511 0.398383	.0.00647035 0.00605783 .0.0430942 0.0430573 .0.00599627 .0.00117909 .0.00717149 .0.00725813 0.00276858 .0.00491306 0.00836377 .0.010313 0.00130315	0.007282 0.006785 0.005142 0.005055 0.005055 0.005055 0.005050 0.005055 0.005055 0.005055 0.005055	-1075 -1050 -1025 9 -1000 -0975 -0950

OneHotEncoding

```
In [41]: from sklearn.preprocessing import LabelEncoder

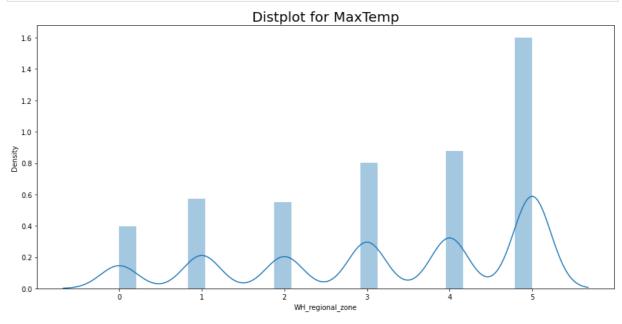
In [18]: le=LabelEncoder()
```

```
In [19]:
           df.head()
Out[19]:
                            WH_Manager_ID
                                             Location_type
                                                           WH_capacity_size
                                                                                    WH_regional_zone
              Ware_house_ID
                                                                              zone
          0
                 WH 100000
                                   EID 50000
                                                     Urban
                                                                       Small
                                                                              West
                                                                                              Zone 6
          1
                 WH_100001
                                   EID_50001
                                                     Rural
                                                                                              Zone 5
                                                                      Large
                                                                             North
          2
                 WH 100002
                                                                                              Zone 2
                                   EID 50002
                                                     Rural
                                                                        Mid
                                                                             South
          3
                 WH 100003
                                   EID_50003
                                                     Rural
                                                                        Mid
                                                                             North
                                                                                              Zone 3
                 WH 100004
                                   EID 50004
                                                                                              Zone 5
          4
                                                     Rural
                                                                      Large North
          5 rows × 24 columns
In [20]:
           df['WH_regional_zone']=le.fit_transform(df['WH_regional_zone'])
In [21]:
           df['WH_regional_zone'].value_counts()
                7376
Out[21]:
                4045
          3
                3708
          1
                2642
          2
                2552
                1827
          Name: WH_regional_zone, dtype: int64
In [22]:
           df.head()
Out[22]:
              Ware_house_ID WH_Manager_ID Location_type WH_capacity_size
                                                                              zone
                                                                                    WH_regional_zone
                                                                                                      nur
          0
                 WH_100000
                                   EID_50000
                                                     Urban
                                                                       Small
                                                                              West
                                                                                                   5
          1
                 WH_100001
                                   EID_50001
                                                     Rural
                                                                       Large
                                                                             North
                                                                                                   4
          2
                                   EID_50002
                                                                             South
                 WH_100002
                                                     Rural
                                                                        Mid
                                                                                                   1
          3
                 WH_100003
                                   EID_50003
                                                     Rural
                                                                        Mid
                                                                             North
                                                                                                   2
          4
                 WH_100004
                                   EID_50004
                                                     Rural
                                                                                                   4
                                                                      Large North
          5 rows × 24 columns
In [23]:
           df.skew()
          WH_regional_zone
                                           -0.544420
Out[23]:
          num refill req 13m
                                           -0.081390
          transport_issue_l1y
                                            1.605424
          Competitor_in_mkt
                                            0.985102
          retail shop num
                                            0.905324
          distributor_num
                                            0.017210
          flood_impacted
                                            2.691308
          flood proof
                                            3.925685
          electric_supply
                                           -0.657167
          dist_from_hub
                                          -0.009042
```

```
1.042478
workers_num
                               0.007485
wh_est_year
storage_issue_reported_13m
                               0.117473
temp_reg_mach
                               0.851244
wh breakdown 13m
                              -0.072809
govt_check_13m
                              -0.357737
product_wg_ton
                               0.336012
```

dtype: float64

```
In [24]:
          plt.figure(figsize=(15,7))
          plt.title('Distplot for MaxTemp', fontsize=20)
          sns.distplot(df['WH_regional_zone'])
          plt.show()
```



We consider Highest curve means -ve skewed -0.54Data is not normal distribution

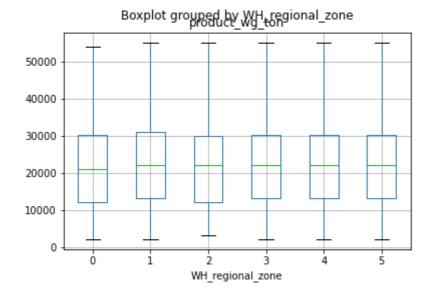
```
In [31]:
          df['WH_regional_zone'].mean()
          3.2474040632054177
Out[31]:
In [32]:
          df['product wg ton'].mean()
          22086.780812641082
Out[32]:
```

Conducting Hypothesis Testing

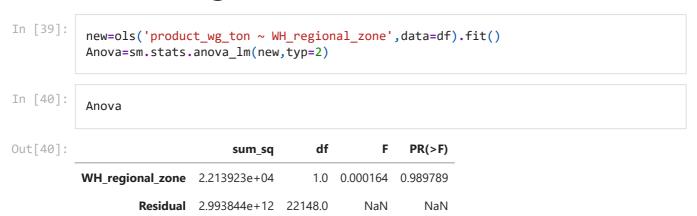
Null Hypothesis = Average product weight shipment equal for all zones

Alternative Hypothesis = Average product weight shipment Not equal for all zones

```
In [35]:
          import statsmodels.api as sm
          from statsmodels.formula.api import ols
```



Conducting Anova Test:



Here P>0.05 [0.98>0.05] so

Null Hypothesis Accepted and Alternative Hypothesis Rejected due to P value not less than 0.005

Average product weight of shipment is equal for all Zones