**proc** **import** datafile='C:\Users\SWAPNIL\Downloads\Documents\AV\markett\Train\_UWu5bXk.csv'

dbms=csv out=market replace;

**run**;

/\* ITEM WEIGHT IMPUTATION \*/

**data** delete;

set market;

if item\_weight=**.** then delete;

**proc** **sort**;

by item\_identifier outlet\_identifier item\_mrp;

**run**;

**proc** **sort** data=market;

by item\_identifier outlet\_identifier item\_mrp;

**run**;

**data** extra\_variable;

set delete;

keep item\_identifier item\_weight;

rename item\_weight=item\_weight2;

**run**;

**data** final;

merge market extra\_variable;

by item\_identifier ;

if item\_weight=**.** then item\_weight=item\_weight2;

**run**;

\*\*\* NEW DATASET \*\*;

**data** new;

set final;

drop item\_weight2;

if item\_weight=**.** then item\_weight=**10**;

**run**;

**data** food;

set new;

if item\_type="Baking Goods" then baking\_goods=**1**;

else baking\_goods=**0**;

if item\_type="Breads" then breads=**1**;

else breads=**0**;

if item\_type="Breakfast" then breakfast=**1**;

else breakfast=**0**;

if item\_type="Canned" then canned=**1**;

else canned=**0**;

if item\_type="Dairy" then dairy=**1**;

else dairy=**0**;

if item\_type="Frozen Foods" then frozen\_foods=**1**;

else frozen\_foods=**0**;

if item\_type="Fruits and Vegetables" then fruits\_vegetables=**1**;

else fruits\_vegetables=**0**;

if item\_type="Hard Drinks" then hard\_drinks=**1**;

else hard\_drinks=**0**;

if item\_type="Health and Hygiene" then health\_hygiene=**1**;

else health\_hygiene=**0**;

if item\_type="Household" then household=**1**;

else household=**0**;

if item\_type="Meat" then meat=**1**;

else meat=**0**;

if item\_type="Seafood" then seafood=**1**;

else seafood=**0**;

if item\_type="Snack Foods" then snack\_foods=**1**;

else snack\_foods=**0**;

if item\_type="Soft Drinks" then soft\_drinks=**1**;

else soft\_drinks=**0**;

if item\_type="Starchy Foods" then starchy\_foods=**1**;

else starchy\_foods=**0**;

if item\_type="Others" then others=**1**;

else others=**0**;

if item\_fat\_content="LF" or item\_fat\_content="Low Fat" or item\_fat\_content="low fat" then fat=**0**;

else fat=**1**;

**proc** **freq**;

table baking\_goods breads breakfast canned dairy frozen\_foods fruits\_vegetables hard\_drinks health\_hygiene

household meat seafood snack\_foods soft\_drinks starchy\_foods others;

**run**;

**data** out010 out013 out017 out018 out019 out027 out035 out045 out046 out049;

set food;

if outlet\_identifier="OUT010" then output out010;

if outlet\_identifier="OUT013" then output out013;

if outlet\_identifier="OUT017" then output out017;

if outlet\_identifier="OUT018" then output out018;

if outlet\_identifier="OUT019" then output out019;

if outlet\_identifier="OUT027" then output out027;

if outlet\_identifier="OUT035" then output out035;

if outlet\_identifier="OUT045" then output out045;

if outlet\_identifier="OUT046" then output out046;

if outlet\_identifier="OUT049" then output out049;

**run**;

**proc** **reg** data=out010;

model item\_outlet\_sales= item\_mrp /vif;

output out=out010 p=pred r=residuals;

**run**;

**proc** **reg** data=out013;

model item\_outlet\_sales= item\_mrp health\_hygiene /vif;

output out=out013 p=pred r=residuals;

**run**;

**proc** **reg** data=out017;

model item\_outlet\_sales= item\_mrp item\_visibility canned/vif;

output out=out017 p=pred r=residuals;

**run**;

**proc** **reg** data=out018;

model item\_outlet\_sales= item\_mrp seafood/vif;

output out=out018 p=pred r=residuals;

**run**;

**proc** **reg** data=out019;

model item\_outlet\_sales= item\_mrp canned /vif;

output out=out019 p=pred r=residuals;

**run**;

**proc** **reg** data=out027;

model item\_outlet\_sales= item\_mrp fruits\_vegetables breakfast /vif;

output out=out027 p=pred r=residuals;

**run**;

**proc** **reg** data=out035;

model item\_outlet\_sales= item\_mrp breakfast seafood starchy\_foods /vif;

output out=out035 p=pred r=residuals;

**run**;

**proc** **reg** data=out045;

model item\_outlet\_sales= item\_mrp fat seafood /vif;

output out=out045 p=pred r=residuals;

**run**;

**proc** **reg** data=out046;

model item\_outlet\_sales= item\_mrp seafood health\_hygiene baking\_goods breads canned /vif;

output out=out046 p=pred r=residuals;

**run**;

**proc** **reg** data=out049;

model item\_outlet\_sales= item\_mrp seafood dairy household /vif;

output out=out049 p=pred r=residuals;

**run**;

**data** mega;

set out010 out013 out017 out018 out019 out027 out035 out045 out046 out049;

**run**;

**proc** **export** data=mega (keep= item\_outlet\_sales pred residuals)

outfile="C:\Users\SWAPNIL\Downloads\Documents\AV\markett\mega.csv"

dbms=csv replace;

**run**;

\*\* TEST DATASET \*\*;

**proc** **import** datafile='C:\Users\SWAPNIL\Downloads\Documents\AV\markett\Test\_u94Q5KV.csv'

dbms=csv out=validation replace;

**run**;

**data** delete2;

set validation;

if item\_weight=**.** then delete;

**proc** **sort**;

by item\_identifier;

**run**;

**proc** **sort** data=validation;

by item\_identifier;

**run**;

**data** new\_variable;

set delete2;

rename item\_weight=item\_weight2;

keep item\_identifier item\_weight;

**run**;

**data** final\_dataset;

merge validation new\_variable;

by item\_identifier ;

if item\_weight=**.** then item\_weight=item\_weight2;

**run**;

\*\*\* NEW DATASET \*\*;

**data** new\_validation;

set final\_dataset;

drop item\_weight2;

if item\_weight=**.** then item\_weight=**10**;

**run**;

**data** store;

set new\_validation;

if item\_fat\_content="LF" or item\_fat\_content="Low Fat" or item\_fat\_content="low fat" then fat=**0**;

else fat=**1**;

if item\_type="Baking Goods" then baking\_goods=**1**;

else baking\_goods=**0**;

if item\_type="Breads" then breads=**1**;

else breads=**0**;

if item\_type="Breakfast" then breakfast=**1**;

else breakfast=**0**;

if item\_type="Canned" then canned=**1**;

else canned=**0**;

if item\_type="Dairy" then dairy=**1**;

else dairy=**0**;

if item\_type="Frozen Foods" then frozen\_foods=**1**;

else frozen\_foods=**0**;

if item\_type="Fruits and Vegetables" then fruits\_vegetables=**1**;

else fruits\_vegetables=**0**;

if item\_type="Hard Drinks" then hard\_drinks=**1**;

else hard\_drinks=**0**;

if item\_type="Health and Hygiene" then health\_hygiene=**1**;

else health\_hygiene=**0**;

if item\_type="Household" then household=**1**;

else household=**0**;

if item\_type="Meat" then meat=**1**;

else meat=**0**;

if item\_type="Seafood" then seafood=**1**;

else seafood=**0**;

if item\_type="Snack Foods" then snack\_foods=**1**;

else snack\_foods=**0**;

if item\_type="Soft Drinks" then soft\_drinks=**1**;

else soft\_drinks=**0**;

if item\_type="Starchy Foods" then starchy\_foods=**1**;

else starchy\_foods=**0**;

if item\_type="Others" then others=**1**;

else others=**0**;

**run**;

**data** predicted;

set store;

if outlet\_identifier="OUT010" then

Item\_Outlet\_Sales= -**27.19247**

+ ( **2.60371** \* Item\_MRP) ;

if outlet\_identifier="OUT013" then

Item\_Outlet\_Sales= -**118.26978**

+ ( **17.20958**\*Item\_MRP ) + ( -**253.89162** \* health\_hygiene) ;

if outlet\_identifier="OUT017" then

Item\_Outlet\_Sales= **35.24207**

+ ( **17.08106**\*Item\_MRP ) + ( -**1553.79026** \* Item\_Visibility ) + ( **259.52563** \* canned ) ;

if outlet\_identifier="OUT018" then

Item\_Outlet\_Sales= -**79.37062**

+ ( **14.59306**\* Item\_MRP) + ( **973.69276** \* seafood ) ;

if outlet\_identifier="OUT019" then

Item\_Outlet\_Sales= **9.14438**

+ ( **2.32098** \* Item\_MRP) + ( **93.68727** \* canned ) ;

if outlet\_identifier="OUT027" then

Item\_Outlet\_Sales= **101.04406**

+ ( **25.26070** \* Item\_MRP) + ( **350.39146**\* fruits\_vegetables )

+ ( **768.27291** \* breakfast ) ;

if outlet\_identifier="OUT035" then

Item\_Outlet\_Sales= -**72.75181**

+ ( **17.37416**\*Item\_MRP ) + ( **683.84031** \* breakfast )

+ ( **780.58931** \* seafood ) + ( **533.18750**\* starchy\_foods) ;

if outlet\_identifier="OUT045" then

Item\_Outlet\_Sales= -**56.32699**

+ ( **15.56536**\* Item\_MRP) + ( **137.08114**\* fat) + ( **779.49798**\* seafood) ;

if outlet\_identifier="OUT046" then

Item\_Outlet\_Sales= **7.26091**

+ ( **15.31491** \* Item\_MRP) + ( **1685.62769** \* seafood) + ( **358.23708**\* health\_hygiene)

+ ( **367.50348**\* baking\_goods) + ( **430.89001** \* breads ) + ( **264.30359** \* canned) ;

if outlet\_identifier="OUT049" then

Item\_Outlet\_Sales= -**60.39474** + ( **17.18973** \* Item\_MRP) + ( -**705.24460** \* seafood )

+ ( -**240.51998**\* dairy) + ( **220.75901**\* household) ;

**run**;

**proc** **export** data=predicted ( keep = Item\_Identifier Outlet\_Identifier Item\_Outlet\_Sales)

outfile="C:\Users\SWAPNIL\Downloads\Documents\AV\markett\predsales.csv"

dbms=csv replace;

**run**;