**/\* FOR THIS FIRST PART, PLEASE REFER TO THE FILE NAMED “RagunathDatga2” \*/**

libname project5 'C:\Users\mxt230013\Desktop\PROJECT5';

**proc** **import** datafile='C:\Users\mxt230013\Desktop\PROJECT5\RagunathData2.csv'

out=project5.RagunathData2

dbms=csv

replace;

**run**;

**proc** **contents** data=project5.RagunathData2;

**run**;

/\* Generate summary statistics for numeric variables grouped by Geography \*/

**proc** **means** data=project5.RagunathData2 n mean std min max;

class Geography;

var \_numeric\_; /\* This specifies that the statistics should be calculated for all numeric variables \*/

output out=geography\_summary; /\* Optionally save the output to a new dataset \*/

**run**;

/\* Summary statistics for selected numeric variables grouped by Geography \*/

**proc** **means** data=project5.RagunathData2 n mean std min max;

class Geography;

var Unit\_Sales Dollar\_Sales Volume\_Sales; /\* Specify the numeric variables of interest \*/

output out=geography\_numeric\_summary; /\* Optionally save the output to a new dataset \*/

**run**;

**/\* TOP BRANDS BY DOLLAR SALES \*/**

**proc** **sql**;

create table brand\_dollarsales as

select Brand, sum(Dollar\_Sales) as Total\_Dollar\_Sales

from project5.RagunathData2

group by Brand

order by Total\_Dollar\_Sales desc;

**quit**;

**proc** **print** data=brand\_dollarsales(obs=**10**) noobs label;

title "Top 10 Brands by Dollar Sales";

var Brand Total\_Dollar\_Sales;

format Total\_Dollar\_Sales dollar20.2;

**run**;

**/\* TOP BRANDS BY UNIT SALES \*/**

**proc** **sql**;

create table brand\_unitsales as

select Brand, sum(Unit\_Sales) as Total\_Unit\_Sales

from project5.RagunathData2

group by Brand

order by Total\_Unit\_Sales desc;

**quit**;

**proc** **print** data=brand\_unitsales(obs=**10**) noobs label;

title "Top 10 Brands by Unit Sales";

var Brand Total\_Unit\_Sales;

**run**;

**/\* THE TOP 4 BRANDS AND THEIR REGIONS MARKET SHARE\*/**

/\* Step 1: Calculate total dollar sales for each brand within each geography \*/

**proc** **sql**;

create table geo\_brand\_dollarsales as

select Geography, Brand, sum(Dollar\_Sales) as Total\_Dollar\_Sales

from project5.RagunathData2

group by Geography, Brand;

**quit**;

/\* Step 2: Calculate total dollar sales for each geography \*/

**proc** **sql**;

create table geo\_totalsales as

select Geography, sum(Total\_Dollar\_Sales) as Geography\_Total\_Sales

from geo\_brand\_dollarsales

group by Geography;

**quit**;

/\* Step 3: Compute the percentages \*/

**proc** **sql**;

create table brand\_dollarsales\_percentage as

select a.Geography, a.Brand, a.Total\_Dollar\_Sales,

(a.Total\_Dollar\_Sales / b.Geography\_Total\_Sales) as Percentage

from geo\_brand\_dollarsales as a

left join geo\_totalsales as b on a.Geography = b.Geography;

**quit**;

/\* Sort by geography and descending total dollar sales to prepare for ranking \*/

**proc** **sort** data=brand\_dollarsales\_percentage out=sorted\_dollarsales;

by Geography descending Total\_Dollar\_Sales;

**run**;

/\* Step 4: Rank the brands within each geography and keep the top 4 \*/

**data** top4\_geo\_brand\_dollarsales;

set sorted\_dollarsales;

by Geography;

if first.Geography or first.Brand then rank = **1**;

else rank + **1**;

if rank <= **4**;

**run**;

/\* Step 5: Display the results \*/

**proc** **print** data=top4\_geo\_brand\_dollarsales noobs;

var Geography Brand Total\_Dollar\_Sales Percentage;

format Total\_Dollar\_Sales dollar20.2 Percentage percent8.2;

title "Top 4 Brands by Dollar Sales Across Each Geography";

**run**;

**/\* VISUALIZATION GRAPHS \*/**

**proc** **sgplot** data=top\_brand\_per\_flavor\_geo;

vbar Geography / response=Total\_Dollar\_Sales group=Flavor groupdisplay=cluster;

xaxis label="Geography";

yaxis label="Total Dollar Sales";

title "Total Dollar Sales of Leading Brands by Flavor Across Geographies";

**run**;

**/\* NORMALIZE FLAVORS \*/**

/\* Step 1: Normalize flavors and brands, filter for specific flavors and their variations \*/

**proc** **sql**;

create table filtered\_normalized\_data as

select

Geography,

/\* Normalize brands \*/

case

when upcase(Brand) like '%MORNINGSTAR%' then 'MORNINGSTAR FARMS'

when upcase(Brand) like '%IMPOSSIBLE%' then 'IMPOSSIBLE'

when upcase(Brand) like '%GARDEIN%' then 'GARDEIN'

else Brand

end as Brand,

/\* Normalize flavors \*/

case

when upcase(Flavor) like '%CHICKEN%' then 'CHICKEN'

when upcase(Flavor) like '%ORIGINAL%' then 'ORIGINAL'

when upcase(Flavor) like '%REGULAR%' then 'REGULAR'

else 'OTHER'

end as Flavor,

Dollar\_Sales,

Unit\_Sales

from project5.RagunathData2

where upcase(Flavor) like '%CHICKEN%'

or upcase(Flavor) like '%ORIGINAL%'

or upcase(Flavor) like '%REGULAR%';

**quit**;

/\* Step 2: Compute the total dollar and unit sales for the normalized brands and flavors within each geography \*/

**proc** **sql**;

create table flavor\_brand\_geo\_sales as

select

Geography,

Flavor,

Brand,

sum(Dollar\_Sales) as Total\_Dollar\_Sales,

sum(Unit\_Sales) as Total\_Unit\_Sales

from filtered\_normalized\_data

where Brand in ('MORNINGSTAR FARMS', 'IMPOSSIBLE', 'GARDEIN')

group by Geography, Flavor, Brand;

**quit**;

**/\* BRAND SHARE AND DOLLAR TO UNIT RATIO FOR EACH FLAVOR ACROSS GEOGRAPHIES \*/**

/\* Step 1: Normalize the data, focusing on the three brands and flavors \*/

/\* This step remains the same as previously provided to normalize brands and filter flavors \*/

/\* Step 2: Compute total dollar and unit sales for each brand, flavor, and geography combination \*/

/\* This step is the same as provided earlier to create the brand\_flavor\_geo\_sales table \*/

/\* Step 3: Calculate total sales for each flavor within each geography to use for calculating percentages \*/

**proc** **sql**;

create table flavor\_geo\_totals as

select Geography, Flavor,

sum(Total\_Dollar\_Sales) as Flavor\_Geo\_Total\_Dollar\_Sales,

sum(Total\_Unit\_Sales) as Flavor\_Geo\_Total\_Unit\_Sales

from flavor\_brand\_geo\_sales

group by Geography, Flavor;

**quit**;

/\* Step 4: Calculate the market share and dollar to unit ratio for each brand within each flavor and geography \*/

**proc** **sql**;

create table brand\_flavor\_geo\_percentage as

select a.Geography, a.Flavor, a.Brand,

a.Total\_Dollar\_Sales, a.Total\_Unit\_Sales,

(a.Total\_Dollar\_Sales / b.Flavor\_Geo\_Total\_Dollar\_Sales) as Brand\_Percentage,

(a.Total\_Dollar\_Sales / a.Total\_Unit\_Sales) as Dollar\_to\_Unit\_Ratio

from flavor\_brand\_geo\_sales as a

left join flavor\_geo\_totals as b

on a.Geography = b.Geography and a.Flavor = b.Flavor;

**quit**;

/\* Step 5: Present the results \*/

/\* Adjust the proc print to show the new Brand\_Percentage and Dollar\_to\_Unit\_Ratio \*/

**proc** **print** data=brand\_flavor\_geo\_percentage noobs label;

var Geography Flavor Brand Total\_Dollar\_Sales Total\_Unit\_Sales Brand\_Percentage Dollar\_to\_Unit\_Ratio;

label Brand = 'Leading\_Brand';

format Total\_Dollar\_Sales dollar20.2 Total\_Unit\_Sales comma20.2 Brand\_Percentage percent8.2 Dollar\_to\_Unit\_Ratio dollar20.2;

title "Brand Share and Dollar to Unit Ratio for Each Flavor Across Geographies";

**run**;

**/\* CONTINUE WITH GARDEIN \*/**

**proc** **sql**;

create table gardein\_flavors as

select \*

from project5.RagunathData2

where upcase(Brand) = 'GARDEIN'

and (upcase(Flavor) = 'ORIGINAL' or

upcase(Flavor) = 'REGULAR' or

upcase(Flavor) = 'CHICKEN');

**quit**;

**proc** **print** data=gardein\_flavors (obs=**10**);

**run**;

**proc** **means** data=gardein\_flavors n mean std min max median;

class Geography Flavor;

var Dollar\_Sales Unit\_Sales;

title "Descriptive Statistics for GARDEIN Brand by Flavor Across Geographies";

**run**;

/\* **SCATTER PLOTS FOR EACH FLAVOR WITHIN A SINGLE GEOGRAPHY** \*/

**proc** **sgplot** data=gardein\_flavors;

where Geography = 'California - Standard'; /\* Replace with your specific geography \*/

scatter x=Unit\_Sales y=Dollar\_Sales / group=Flavor markerattrs=(symbol=CircleFilled);

xaxis label="Unit Sales";

yaxis label="Dollar Sales";

title "Scatter Plot of Unit Sales vs Dollar Sales for GARDEIN Brand";

**run**;

**/\* CORRECTIONS SCATTER PLOTS\*/**

**proc** **sgplot** data=WORK.GARDEIN\_FLAVORS;

where Geography='California - Standard';

scatter x=Unit\_Sales y=Dollar\_Sales / group=Flavor markerattrs=(symbol=CircleFilled);

reg x=Unit\_Sales y=Dollar\_Sales / group=Flavor; /\* Corrects the fit line \*/

xaxis label="Unit Sales" grid;

yaxis label="Dollar Sales" grid;

title "Enhanced Scatter Plot of Unit Sales vs Dollar Sales for GARDEIN Brand by Flavor";

keylegend / title="Flavor";

**run**;

**/\* FLAVOR FREQUENT OR COUNTS \*/**

**proc** **sql** noprint;

create table WORK.GARDEIN\_FLAVORS as

select \*

from project5.RagunathData2

where upcase(Brand) = 'GARDEIN' and

(upcase(Flavor) contains 'CHICKEN' or

upcase(Flavor) contains 'ORIGINAL' or

upcase(Flavor) contains 'REGULAR');

**quit**;

**proc** **freq** data=WORK.GARDEIN\_FLAVORS;

tables Flavor / nocum nopercent;

**run**;

**/\* SCATTER PLOT OF Unit Sales VS. DOLLAR SALES \*/**

**proc** **sgplot** data=WORK.GARDEIN\_FLAVORS;

scatter x=Unit\_Sales y=Dollar\_Sales / group=Flavor transparency=**0.5**;

reg x=Unit\_Sales y=Dollar\_Sales / group=Flavor;

xaxis label="Unit Sales";

yaxis label="Dollar Sales";

title "Scatter Plot of Unit Sales vs Dollar Sales for All GARDEIN Flavor Variations";

keylegend / title="Flavor";

**run**;

/\* Histograms for Unit Sales \*/

**proc** **sgplot** data=WORK.GARDEIN\_FLAVORS;

histogram Unit\_Sales / binwidth = **500**; /\* Adjust the binwidth as appropriate for your data \*/

density Unit\_Sales / type = kernel;

density Unit\_Sales / type = normal;

title 'Unit Sales Distribution for GARDEIN Brand';

**run**;

**/\*\* For this second part, please refer to the clean data named “**RagunathData4**”** **\*//**

libname project6 'C:\Users\mxt230013\Desktop\PROJECT6';

**proc** **import** datafile='C:\Users\mxt230013\Desktop\PROJECT6\RagunathData4.csv'

out=project6.RagunathData4

dbms=csv

replace;

**run**;

**proc** **contents** data=project6.RagunathData4;

**run**;

**proc** **freq** data=project6.RagunathData4;

where Flavor in ('CHICKEN', 'REGULAR', 'ORIGINAL');

tables Flavor / nocum nopercent;

**run**;

/\* ANOVA TEST FOR THE GEOGRAPHIC REGIONS \*/

**proc** **anova** data=project6.RagunathData4;

class Geography;

model Price\_per\_Unit = Geography;

where Brand='GARDEIN' and Flavor='CHICKEN';

title The ANOVA Procedure;

**run**;

/\* MODEL PREDICTORS SELECTION PROCESS \*/

**proc** **corr** data=project6.RagunathData4;

where Brand = 'GARDEIN' and Flavor = 'CHICKEN';

var Dollar\_Sales Price\_per\_Unit Volume\_Sales Unit\_sales Year;

**run**;

/\* Histogram of Unit\_Sales for GARDEIN CHICKEN Flavor \*/

ods graphics on;

**proc** **sgplot** data=project6.RagunathData4;

where Brand='GARDEIN' and Flavor='CHICKEN';

histogram Unit\_Sales / scale=percent;

density Unit\_Sales / type=kernel lineattrs=(color=blue thickness=**2**); /\* Kernel curve in blue \*/

density Unit\_Sales / type=normal lineattrs=(color=red thickness=**2**); /\* Normal curve in red \*/

keylegend / location=inside position=topright across=**1**; /\* Add a legend inside the plot \*/

title 'Histogram of Unit\_Sales for GARDEIN CHICKEN Flavor';

**run**;

ods graphics off;

ods graphics on;

/\*HISTOGRAM OF Price\_per\_Unit FOR GARDEIN CHICKEN FALAVOR \*\*/

**proc** **sgplot** data=project6.RagunathData4;

where Brand='GARDEIN' and Flavor='CHICKEN';

histogram Price\_per\_Unit / scale=percent;

density Price\_per\_Unit / type=kernel lineattrs=(color=blue thickness=**2**); /\* Kernel curve in blue \*/

density Price\_per\_Unit / type=normal lineattrs=(color=red thickness=**2**); /\* Normal curve in red \*/

keylegend / location=inside position=topright across=**1**; /\* Add a legend inside the plot \*/

title 'Histogram of Price\_per\_Unit for GARDEIN CHICKEN Flavor';

**run**;

ods graphics off;

/\*Histogram of ALL Price per Unit for GARDEIN REGULAR FLAVOR\*/

ods graphics on;

**proc** **sgplot** data=project6.RagunathData4;

where Brand='GARDEIN' and Flavor='REGULAR';

histogram Price\_per\_Unit / binstart=**0.5** binwidth=**0.1** scale=percent;

density Price\_per\_Unit / type=kernel lineattrs=(color=blue); /\* Kernel curve in blue \*/

density Price\_per\_Unit / type=normal lineattrs=(color=red); /\* Normal curve in red \*/

keylegend / location=inside position=topright across=**1**; /\* Optional: Add a legend inside the plot \*/

title "Histogram of ALL Price per Unit for GARDEIN REGULAR FLAVOR";

**run**;

/\* HISTOGRAM PLOT OF LOW PRICES FOR GARDEIN REGULAR FLAVOR\*/

ods graphics on;

**proc** **sgplot** data=project6.RagunathData4;

where Brand='GARDEIN' and Flavor='REGULAR' and Price\_per\_Unit < **6.8**;

histogram Price\_per\_Unit / binstart=**0.5** binwidth=**0.1** scale=percent;

density Price\_per\_Unit / type=kernel lineattrs=(color=blue); /\* Kernel curve in blue \*/

density Price\_per\_Unit / type=normal lineattrs=(color=red); /\* Normal curve in red \*/

keylegend / location=inside position=topright across=**1**; /\* Optional: Add a legend inside the plot \*/

title "Histogram of Lower Price per Unit for GARDEIN REGULAR FLAVOR";

**run**;

/\* MODEL SELECTION PROCESS \*/

ods graphics on;

**proc** **reg** data=project6.RagunathData4 plots=all;

model Dollar\_Sales = Unit\_Sales ACV\_Weighted\_Distribution;

where upcase(Flavor) = 'ORIGINAL' and upcase(Brand) = 'GARDEIN';

title "Regression Analysis with Interaction Terms for GARDEIN Brandfs ORIGINAL Flavor";

**run**;

ods graphics off;

ods graphics on;

**proc** **reg** data=project6.RagunathData4 plots=all;

model Dollar\_Sales =Volume\_Sales Price\_per\_Unit Unit\_Sales Year;

where upcase(Flavor) = 'REGULAR' and upcase(Brand) = 'GARDEIN' and Price\_Range= **0**;

title "Regression Analysis with Interaction Terms for GARDEIN Brand?fs REGULAR Flavor";

**run**;

ods graphics off;

/\* REGRESSION MODEL FOR GARDEIN \*/

ods graphics on;

**proc** **reg** data=project6.RagunathData4 plots=all;

model Dollar\_Sales =Price\_per\_Unit Volume\_Sales Year;

where upcase(Flavor) = 'REGULAR' and upcase(Brand) = 'GARDEIN' and Price\_Range= **0**;

title "Regression Analysis with for GARDEIN Brand?s REGULAR Flavor";

**run**;

ods graphics off;