**Scanner data code**

/\* Step 1: Inspect the Raw Dataset \*/

proc contents data=work.import1;

title "Dataset Structure: Scanner Data (Raw)";

run;

proc print data=work.import1(obs=10);

title "Preview of Raw Scanner Data";

run;

/\* Step 2: Fix Date Format and Rename Columns \*/

data work.scanner\_cleaned;

set work.import1;

/\* Correct the Date format \*/

if \_n\_ = 1 then do;

informat Date ddmmyy10.;

format Date yymmdd10.;

end;

/\* Rename columns for clarity \*/

retain Date Customer\_ID Transaction\_ID SKU\_Category SKU Quantity Sales\_Amount;

/\* Drop unnecessary columns (e.g., VAR1) \*/

drop VAR1;

run;

/\* Step 3: Validate the Cleaned Dataset \*/

proc contents data=work.scanner\_cleaned;

title "Dataset Structure: Scanner Data (Cleaned)";

run;

proc print data=work.scanner\_cleaned(obs=10);

title "Preview of Cleaned Scanner Data";

run;

/\* Step 4: Handle Missing Values \*/

/\* Check for missing values \*/

proc means data=work.scanner\_cleaned n nmiss;

var Quantity Sales\_Amount;

run;

/\* Impute missing numeric values with the mean \*/

proc stdize data=work.scanner\_cleaned reponly missing=mean out=work.scanner\_final;

var Quantity Sales\_Amount;

run;

/\* Step 5: Validate the Final Dataset \*/

proc contents data=work.scanner\_final;

title "Dataset Structure: Scanner Data (Final)";

run;

proc print data=work.scanner\_final(obs=10);

title "Preview of Final Scanner Data";

run;

/\* Step 6: Summary Statistics \*/

proc means data=work.scanner\_final mean std min max n;

var Quantity Sales\_Amount;

title "Summary Statistics for Scanner Data";

run;

/\* Step 7: Top 10 SKU Categories by Sales \*/

proc sql;

create table sales\_by\_category as

select SKU\_Category,

sum(Sales\_Amount) as Total\_Sales\_Amount

from work.scanner\_final

group by SKU\_Category

order by Total\_Sales\_Amount desc;

quit;

data top\_10\_categories;

set sales\_by\_category;

if \_N\_ <= 10; /\* Keep only the top 10 rows \*/

run;

proc sgplot data=top\_10\_categories;

vbar SKU\_Category / response=Total\_Sales\_Amount datalabel dataskin=pressed;

xaxis label="SKU Category";

yaxis label="Total Sales (USD)";

title "Top 10 SKU Categories by Sales";

run;

/\* Step 8: Top 10 Products by Sales \*/

proc sql;

create table sales\_by\_sku as

select SKU,

sum(Sales\_Amount) as Total\_Sales\_Amount

from work.scanner\_final

group by SKU

order by Total\_Sales\_Amount desc;

quit;

data top\_10\_skus;

set sales\_by\_sku;

if \_N\_ <= 10; /\* Keep only the top 10 rows \*/

run;

proc sgplot data=top\_10\_skus;

vbar SKU / response=Total\_Sales\_Amount datalabel dataskin=pressed;

xaxis label="Product SKU";

yaxis label="Total Sales (USD)";

title "Top 10 Products by Sales";

run;

/\* Step 9: Correlation Analysis Between Quantity and Sales \*/

proc corr data=work.scanner\_final;

var Quantity Sales\_Amount;

title "The CORR Procedure";

run;

/\* Step 10: Summary Statistics for Correlation Variables \*/

proc means data=work.scanner\_final mean std sum min max;

var Quantity Sales\_Amount;

title "Summary Statistics for Quantity and Sales Amount";

run;