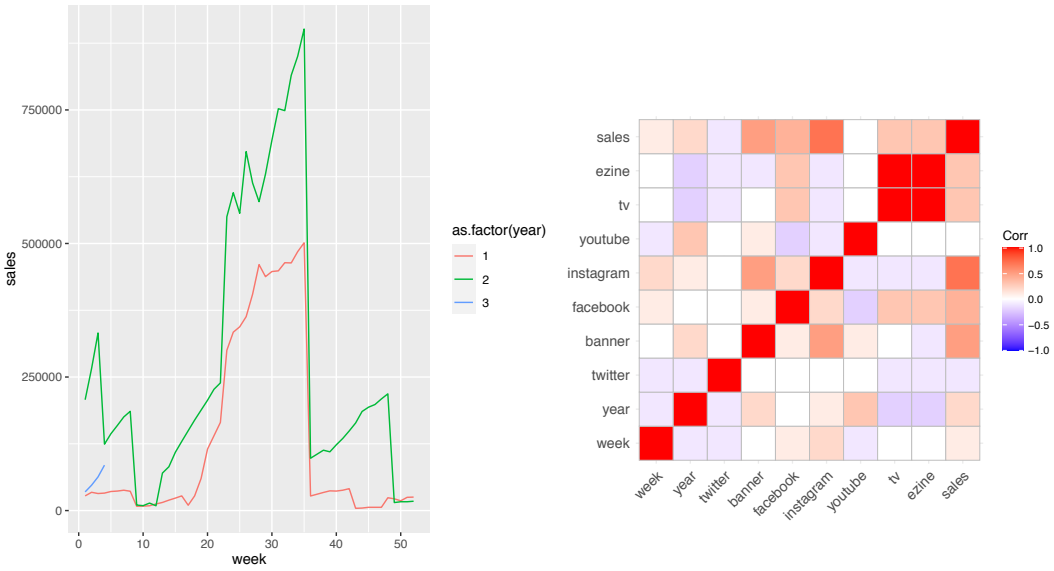
Assignment 3 – Model Answers

Problem 1

Running the R-script gives you the following graph of sales by week and the following correlation plot.



Based on data from Years 1 and 2, we can see a clear trend in higher sales around weeks 20-35. Data from Years 2 and 3 suggests that sales may also increase at the start of the year, although this is speculative since we have limited data from Year 3.

The correlation matrix tells us that TV and Ezine expenditure are highly correlated. We can verify this using the cor() function:

```
cor(assignment3$ezine,assignment3$tv)
[1] 0.9718452
```

There aren't many strong negative correlations. One example of negative correlation is facebook and youtube expenditure:

```
cor(assignment3$facebook,assignment3$youtube)
[1] -0.1539306
```

Problem 2

Running the model gives us the following output:

```
lm1 <- lm(formula = sales ~ twitter + banner + facebook +
instagram +

 youtube + tv + ezine, data = assignment3)

 Residuals:
     Min       1Q   Median       3Q      Max
 -344203  -62349  -26096   38140  519366

 Coefficients:
               Estimate Std. Error t value Pr(>|t|)
 (Intercept) 38665.0177 19397.9674   1.993  0.04896 *
 twitter        -0.0863     1.7940  -0.048  0.96173
 banner         37.5130     8.3657   4.484 1.96e-05 ***
 facebook       72.7548    23.8287   3.053  0.00290 **
 instagram      73.4099    10.0225   7.325 6.24e-11 ***
 youtube         2.6594     8.9153   0.298  0.76609
 tv            -12.1015     6.7824  -1.784  0.07742 .
 ezine          21.1405     6.8838   3.071  0.00275 **
 ---
 Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Residual standard error: 133900 on 100 degrees of freedom
 Multiple R-squared:  0.6778,  Adjusted R-squared:  0.6552
 F-statistic: 30.05 on 7 and 100 DF,  p-value: < 2.2e-16
```

Our model explains roughly 65.5% of variation in sales revenue (see adjusted R-squared). Banner, Facebook, Instagram, and ezine ad expenditure affect sales revenue positively. Of these, Instagram expenditure has the largest effect, increasing revenue by ~73 EUR for each Euro spent. TV ad expenditure appears to have a negative effect on sales revenue, although the effect is only marginally significant ($p>0.05$). Based on these results, we can speculate that Facebook and Instagram advertising yields the greatest ROI in terms of revenue.

Problems 3 and 4 – improving your model

As we saw in Problem 1, our sales data does seem to be affected by seasonality. There are many ways to code the seasonality variable. For example, you can eyeball which weeks exhibit a clear peak based on the graph, or you can compare weekly sales to median or average annual sales. In the model below, I coded my seasonality variable to equal 1 during weeks 22-35.

```
lm2 <- lm(formula = sales ~ twitter + banner + facebook + instagram +
youtube + tv + ezine + season, data = assignment3)

Residuals:
    Min      1Q  Median      3Q     Max
-272612  -41491  -13032   48882  229637

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.378e+04  1.192e+04   4.510 1.78e-05 ***
twitter     -2.127e-01  1.097e+00  -0.194 0.846752
banner       1.598e+01  5.380e+00   2.970 0.003740 **
facebook     5.950e+00  1.546e+01   0.385 0.701153
instagram    3.954e+01  6.664e+00   5.934 4.38e-08 ***
youtube      2.027e+01  5.620e+00   3.607 0.000487 ***
tv          -8.456e+00  4.159e+00  -2.033 0.044682 *
ezine        7.237e+00  4.345e+00   1.665 0.099005 .
season       3.859e+05  2.975e+04  12.970  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 81930 on 99 degrees of freedom
Multiple R-squared:  0.8806,  Adjusted R-squared:  0.871
F-statistic: 91.28 on 8 and 99 DF,  p-value: < 2.2e-16
```

The model has improved in terms of R-squared, from 0.655 to 0.871. There are also some changes to the significance of the independent variables and the sizes of the coefficients. This suggests that our previous model suffered from omitted variable bias due to seasonality.

We can try to further improve the R-squared of the model by removing some variables. Removing Twitter and Facebook ad expenditure gives us the following model, with a marginally higher R-squared:

```
lm3 <- lm(formula = sales ~ banner + instagram + youtube + tv + ezine +
season, data = assignment3)

Residuals:
    Min      1Q  Median      3Q     Max
-277648  -41902  -14368   48110  230171

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  54241.360  11476.314   4.726 7.41e-06 ***
banner          15.910      5.326   2.987 0.003534 **
instagram       39.655      6.597   6.011 2.95e-08 ***
youtube         20.159      5.550   3.632 0.000443 ***
tv              -8.232      4.079  -2.018 0.046215 *
ezine            7.048      4.271   1.650 0.101991
season      389669.646  27799.137  14.017  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 81190 on 101 degrees of freedom
Multiple R-squared:  0.8804,  Adjusted R-squared:  0.8733
F-statistic: 123.9 on 6 and 101 DF,  p-value: < 2.2e-16
```

Problem 5 – Calculating MROI

We can calculate return on marketing investment by plugging the coefficients into the MROI formula as follows:

Write the coefficients from the linear model (in this case, 'lm3') to the object 'coefficients':

```
> coefficients<-lm3$coefficients
```

Calcalte MROI (in %) from investing 20 000 Euros into each channel:

```
> mroi<-((coefficients*20000*0.06)-20000)/20000*100
> mroi
```

```
  names                x
  <chr>            <dbl>
1 (Intercept)  325348.
2 banner           -4.54
3 instagram       138.
4 youtube          21.0
5 tv             -149.
6 ezine            -57.7
7 season      2337918.
```

NB: you will likely have different values for MROI if you coded your seasonality variable differently or ran a different model.

Problem 6

For the purpose of the question, we only need to run a model that includes spend on banner and Instagram advertising, as well as their interaction:

```
> lm7<-lm(sales
        ~banner
        +instagram
        +instagram*banner, data=assignment2)

> tidy(lm7)
```

```
  term                 estimate   std.error statistic       p.value
  <chr>                   <dbl>       <dbl>     <dbl>         <dbl>
1 (Intercept)      100724.      18931.           5.32 0.000000598
2 banner               54.1        12.7          4.27 0.0000428
3 instagram            97.8        15.0          6.53 0.00000000244
4 banner:instagram     -0.00657     0.00287     -2.29 0.0238
```

Plug in the appropriate levels of spending into the model equation as follows:

$Sales = b_0 + b_1\,Banner + b_2\,Instagram + b_3\,(Banner * Instagram)$

$Sales = 100\,724 + 54.1 * 2\,000 + 97.8 * 5\,000 + (-0.00657 * (2\,000 * 5\,000)) =$

$= \underline{632\,224}$

The total sales resulting from spending 5000 EUR on Instagram and 2000 EUR on banner advertising is thus 632 224 Euros. According to this model, without Instagram advertising, sales revenue would 208 924 Euros.