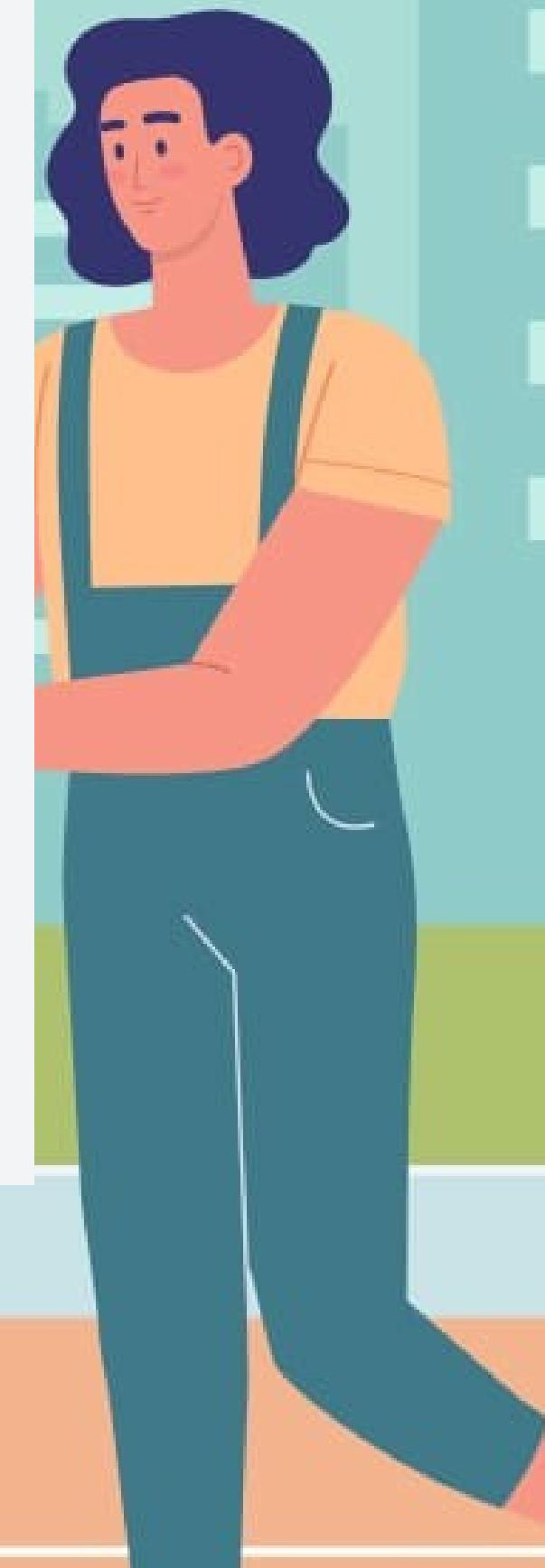


ANALYSING BIKE RENTALS

**DETERMINING THE NUMBER OF
BIKE RENTALS**

OPERATIONS ANALYTICS

**GROUP 4: RUBEN CUESTA, KRISHAN KANT, SARAH RAUBENHEIMER,
JOSEPHINE LELTZ, LUIS GONZALEZ**



CURRENT CHALLENGES

Changing weather conditions lead to unpredictable demand



High inventory costs due to unused bikes



Lost sales due to high demand and bike shortage



ADDRESSING THE CHALLENGES WITH DATA

01

Preparing and exploring data which presents bike rental figures and daily weather conditions

02

Creating models that predict the number of bike rentals which will happen each day (Linear Regression & Neural Nets)

03

Fine-tuning models and deriving business insights from predictions (e.g., feature importance and accuracy)



MODELS AND FINDINGS

Model	Explainability (R2)*	Insights
1. Linear Regression (all variables**)	55.36%	Most significant predictor variables (Date & Weather)
2. Linear Regression (only significant weather variables)	65.06%	Outcome variable prediction
3. Linear Regression*** (only significant variables)	73.41%	Demand prediction (best model to predict bike demand)
4. Neural Network (only significant variables)	60.87%	Black Box model + high computational power needed (not explainable nor scalable)

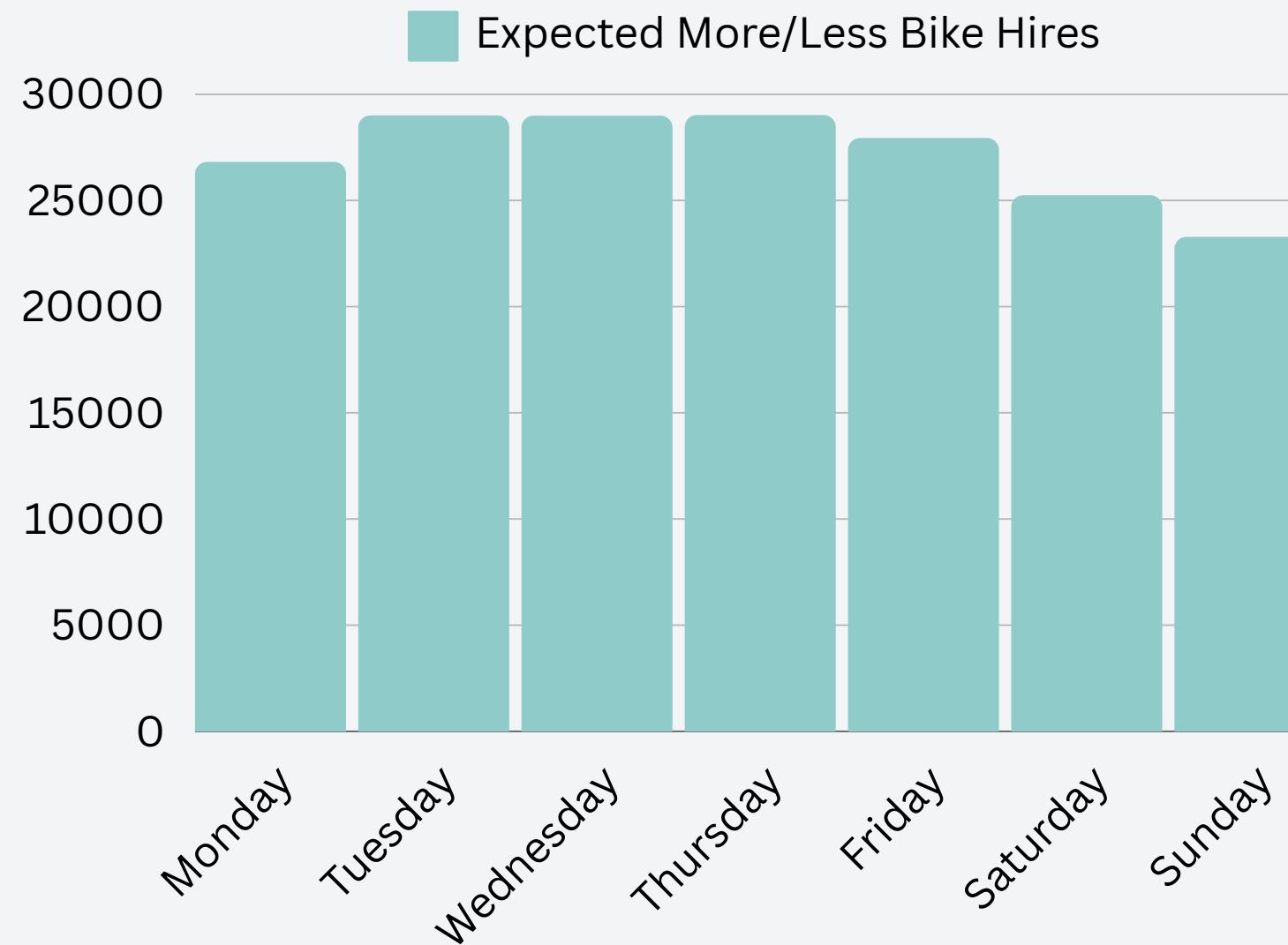
* R2 taken from the testing set (training R2 available in the appendix)

** Reducing the amount of variables will reduce computing power (increasing efficiency) and less data gathered reduces costs

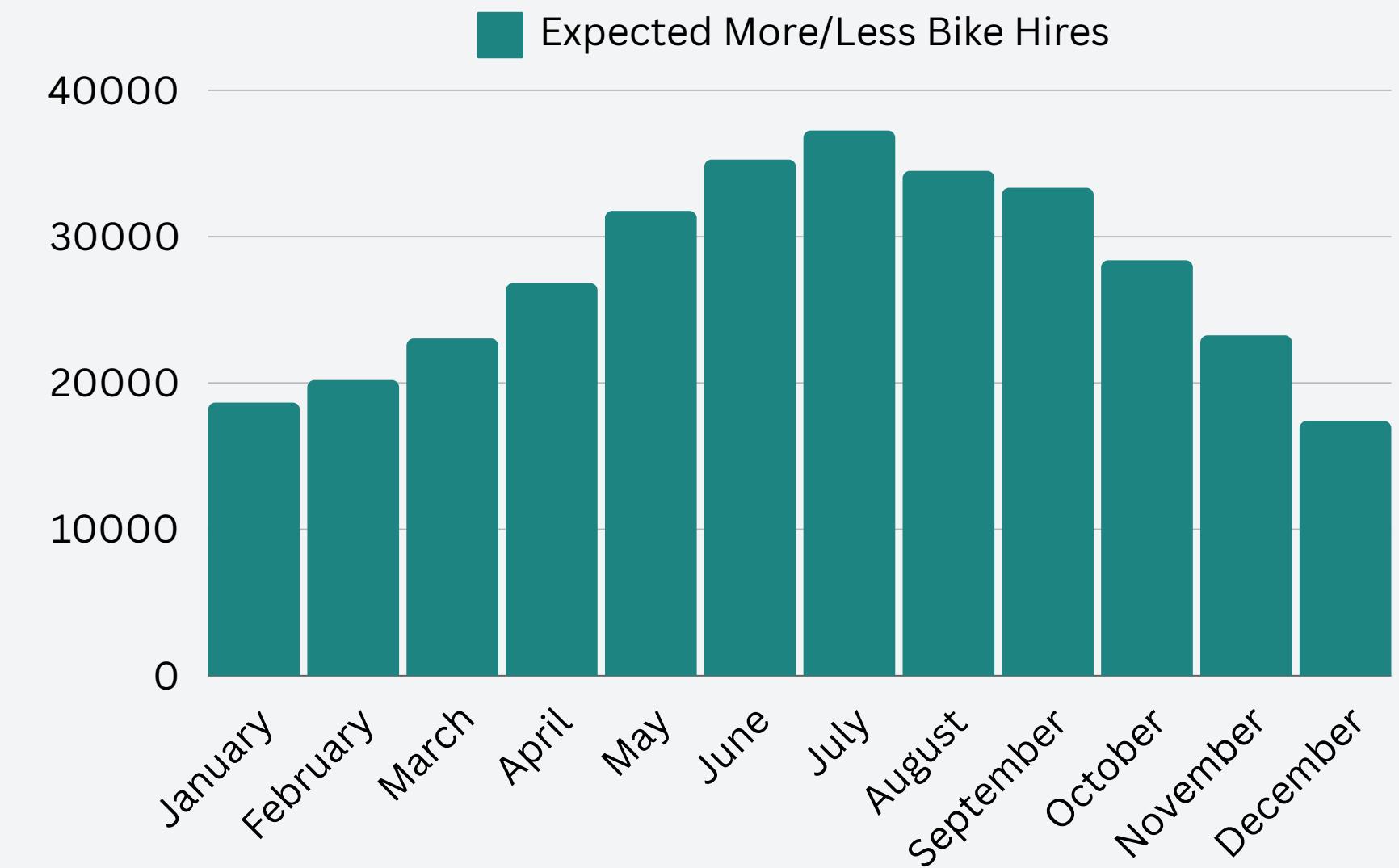
*** Random sampling technique applied

SEASONALITY - TIMING AFFECTS DEMAND

Day of Week



Month



- Wednesday is the best day for bike hires, with Tuesdays and Thursdays closely following.
- Saturdays and Sundays have less bike hires: most customers use bikes to get to work

- Summer and Autumn months are the best months for bike hires, with July being the best performing month.
- During December, bike hires are the lowest compared to the other months.

HOW DOES WEATHER AFFECT DEMAND?

MODEL: LINEAR REGRESSION WITH ONLY SIGNIFICANT VARIABLES

POSITIVELY AFFECTING BIKE RENTALS

AIR PRESSURE

- INCREASE BY 1 --> 7 MORE RENTALS

SUNSHINE

- INCREASE BY 1 --> 49 MORE RENTALS

MEAN TEMPERATURE

- INCREASE BY 1 --> **1311 MORE RENTALS**

NEGATIVELY AFFECTING BIKE RENTALS

HUMIDITY

- INCREASE BY 1 --> **83 LESS RENTALS**

PRECIPITATION

- INCREASE BY 1 --> 33 LESS RENTALS

RADIATION

- INCREASE BY 1 --> 10 LESS RENTALS

BUSINESS RECOMMENDATIONS

Best Prediction Model

Use **Linear Regression** to predict bike rentals (yields best explainability and lowest error terms)

Future models:

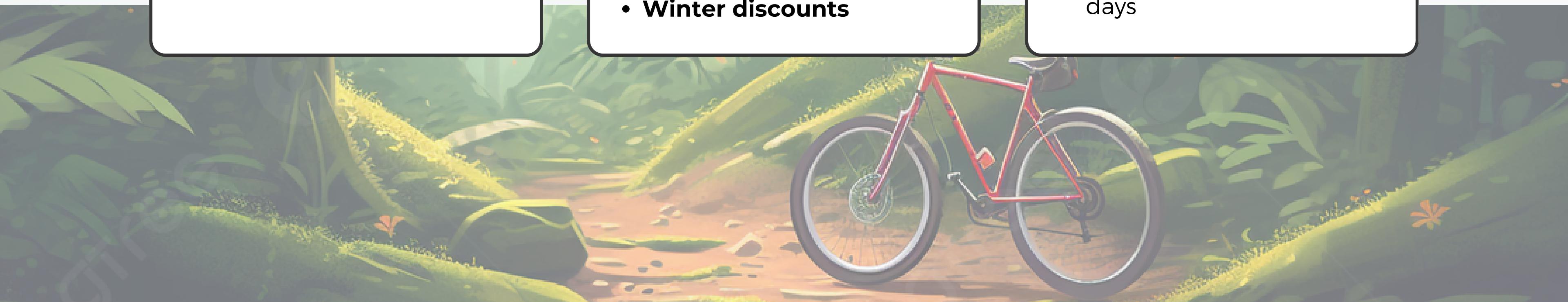
- Customer **demographics**
- Repeat customers

Influence of time

- Have more bikes available on **Tuesdays, Wednesdays & Thursdays**
- Increase capacity from **May** to **November**
- Promote leisure cycling with **weekend packages**
- **Winter discounts**

Influence of Weather

- Incorporate weather forecasts into **capacity management**
- Implement dynamic pricing structure to regulate changing demand
- Provide rain coats on rainy days



APPENDIX STRUCTURE

- Models and findings: From chronological sample
 - Models and error metrics from validation set and chronological sampling
 - Linear regression model 1
 - Linear regression model 2
 - Linear regression model 3
 - Neural Network 1
 - Neural Network 2
- Models and findings From Random sample
 - Logistic regression: Random sampling and all variables
 - Random sampling and only significant variables
- Link to GitHub Code

APPENDIX: MODELS AND FINDINGS: FROM CHRONOLOGICAL SAMPLE

Model	Training R2	Testing R2	Insights
1. Linear Regression (all variables)	76.20%	55.36%	Most significant predictor variables (Date & Weather)
2. Linear Regression (only significant variables)	72.17%	55.22%	Demand prediction (best model to predict bike demand)
3. Linear Regression (only significant weather variables)	63.06%	65.06%	Outcome variable prediction
4. Neural Network (only significant variables)	71.86%	60.87%	Improved outcome prediction

APPENDIX: MODELS AND ERROR METRICS FROM VALIDATION SET AND CHRONOLOGICAL SAMPLING

Model	R2	RMSE	MAPE	MAE	MSE
Linear regression (all variables)	0.5535532	7292.3	0.2082214	5500.301	53177646
Linear regression (significant variables)	0.5521775	7290.713	0.2091074	5508.786	53154492
Linear regression (significant weather variables)	0.650618	7795.286	0.2069615	6328.261	60766490
Neural network* (all numerical variables)	0.1709133	0.2296976	0.4568433	0.18582	0.05276098
Neural network* (significant weather variables)	0.6086737	0.127202	0.2654279	0.1027758	0.01618034

* the data for the neural network has been normalised

APPENDIX: LINEAR REGRESSION MODEL 1

All variables used

Insignificant variables:

- Monday compared to Friday
- January, February, December, March compared to April
- Week
- Cloud cover
- Snow depth
- Radiation
- Maximum temperature

	Training	Testing
R2	0.7620383	0.5535532
RMSE	4542.179	7292.3
MAPE	0.1722726	0.2082214
MAE	3341.411	5500.301
MSE	20631392	53177646

```

Residuals:
    Min      1Q Median     3Q    Max
-23638 -2410   356  2716 31189

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.183e+06 6.325e+04 -18.700 < 2e-16 ***
year         5.717e+02 3.106e+01  18.404 < 2e-16 ***
wdayMon     -5.740e+02 2.988e+02 -1.921 0.054810 .
wdaySat     -4.350e+03 2.982e+02 -14.586 < 2e-16 ***
wdaySun     -5.806e+03 2.985e+02 -19.451 < 2e-16 ***
wdayThu      1.124e+03 2.995e+02  3.753 0.000178 ***
wdayTue      1.221e+03 2.987e+02  4.088 4.46e-05 ***
wdayWed      1.344e+03 2.990e+02  4.496 7.16e-06 ***
monthAug     4.278e+03 5.971e+02  7.165 9.58e-13 ***
monthDec     -1.370e+03 9.325e+02 -1.470 0.141760
monthFeb     -3.530e+02 5.502e+02 -0.642 0.521176
monthJan     -4.012e+02 6.064e+02 -0.662 0.508247
monthJul     5.242e+03 5.627e+02  9.316 < 2e-16 ***
monthJun     4.553e+03 4.981e+02  9.140 < 2e-16 ***
monthMar     -6.483e+02 4.299e+02 -1.508 0.131642
monthMay      3.417e+03 4.302e+02  7.942 2.71e-15 ***
monthNov     2.866e+03 8.678e+02  3.303 0.000968 ***
monthOct      4.204e+03 7.719e+02  5.446 5.53e-08 ***
monthSep      4.447e+03 6.729e+02  6.609 4.51e-11 ***
week        -1.535e+01 2.191e+01 -0.701 0.483611
cloud_cover   6.198e+01 6.682e+01  0.928 0.353724
humidity     -7.033e+01 1.308e+01 -5.375 8.19e-08 ***
pressure      5.031e+00 8.378e-01  6.004 2.13e-09 ***
radiation    -3.398e+00 5.013e+00 -0.678 0.497946
precipitation -3.373e+01 2.358e+00 -14.303 < 2e-16 ***
snow_depth    -5.153e+02 3.004e+02 -1.715 0.086417 .
sunshine       4.462e+01 8.469e+00  5.268 1.47e-07 ***
mean_temp     1.223e+03 8.991e+01  13.606 < 2e-16 ***
min_temp      -6.194e+02 7.380e+01 -8.393 < 2e-16 ***
max_temp      -2.083e+01 3.439e+01 -0.606 0.544728
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4563 on 3243 degrees of freedom
Multiple R-squared:  0.762,    Adjusted R-squared:  0.7599
F-statistic: 358.1 on 29 and 3243 DF,  p-value: < 2.2e-16

```

APPENDIX: LINEAR REGRESSION MODEL 2

Only significant variables included:

	Training	Testing
R2	0.7616849	0.5521775
RMSE	4545.551	7290.713
MAPE	0.1725684	0.2091074
MAE	3341.875	5508.786
MSE	20662030	53154492

- 75.99% of number bikes hired can be explained by model 2
- Error metrics are high because data is not normalised

Residuals:					
	Min	1Q	Median	3Q	Max
	-23788.6	-2421.9	348.6	2720.7	31130.7
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.194e+06	6.229e+04	-19.174	< 2e-16	***
year	5.772e+02	3.061e+01	18.857	< 2e-16	***
wdayMon	-5.880e+02	2.986e+02	-1.969	0.049007	*
wdaySat	-4.352e+03	2.979e+02	-14.610	< 2e-16	***
wdaySun	-5.821e+03	2.983e+02	-19.512	< 2e-16	***
wdayThu	1.129e+03	2.992e+02	3.774	0.000164	***
wdayTue	1.216e+03	2.985e+02	4.074	4.73e-05	***
wdayWed	1.345e+03	2.987e+02	4.503	6.92e-06	***
monthAug	3.888e+03	4.559e+02	8.528	< 2e-16	***
monthDec	-1.669e+03	4.177e+02	-3.995	6.61e-05	***
monthFeb	-1.823e+01	4.164e+02	-0.044	0.965078	
monthJan	2.065e+01	4.219e+02	0.049	0.960955	
monthJul	4.869e+03	4.705e+02	10.350	< 2e-16	***
monthJun	4.260e+03	4.379e+02	9.727	< 2e-16	***
monthMar	-4.517e+02	3.877e+02	-1.165	0.244064	
monthMay	3.256e+03	4.074e+02	7.992	1.83e-15	***
monthNov	2.634e+03	4.121e+02	6.390	1.89e-10	***
monthOct	3.960e+03	4.099e+02	9.661	< 2e-16	***
monthSep	4.131e+03	4.313e+02	9.578	< 2e-16	***
humidity	-6.788e+01	1.274e+01	-5.329	1.06e-07	***
pressure	5.055e+00	8.290e-01	6.098	1.20e-09	***
precipitation	-3.353e+01	2.353e+00	-14.251	< 2e-16	***
sunshine	3.762e+01	3.632e+00	10.358	< 2e-16	***
mean_temp	1.176e+03	7.094e+01	16.580	< 2e-16	***
min_temp	-5.811e+02	6.662e+01	-8.723	< 2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
Residual standard error: 4563 on 3248 degrees of freedom					
Multiple R-squared: 0.7617, Adjusted R-squared: 0.7599					
F-statistic: 432.5 on 24 and 3248 DF, p-value: < 2.2e-16					

APPENDIX: LINEAR REGRESSION MODEL 3

Only significant **weather** variables included:

	Training	Testing
R2	0.6305694	0.650618
RMSE	5659.487	7795.286
MAPE	0.2438269	0.2069615
MAE	4527.838	6328.261
MSE	32029798	60766490

```
Residuals:
    Min      1Q Median     3Q    Max
-20613 -3625    527  4021 34488

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept) -34047.331 10487.287 -3.247 0.00118 **
humidity     -60.292   14.455 -4.171 3.11e-05 ***
pressure      4.960    1.020  4.863 1.21e-06 ***
precipitation -32.661   2.901 -11.257 < 2e-16 ***
sunshine       35.981   4.469  8.051 1.14e-15 ***
mean_temp     1612.054  80.635 19.992 < 2e-16 ***
min_temp     -694.117  81.537 -8.513 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5666 on 3266 degrees of freedom
Multiple R-squared:  0.6306,    Adjusted R-squared:  0.6299
F-statistic: 929.1 on 6 and 3266 DF,  p-value: < 2.2e-16
```

- 62.99% of number bikes hired can be explained by model 2
- Error metrics are higher than model including weather variables
- For the analysis weather variables will be included

APPENDIX: NEURAL NETWORK 1

Model 1: All numeric features

Try to improve error from linear regression

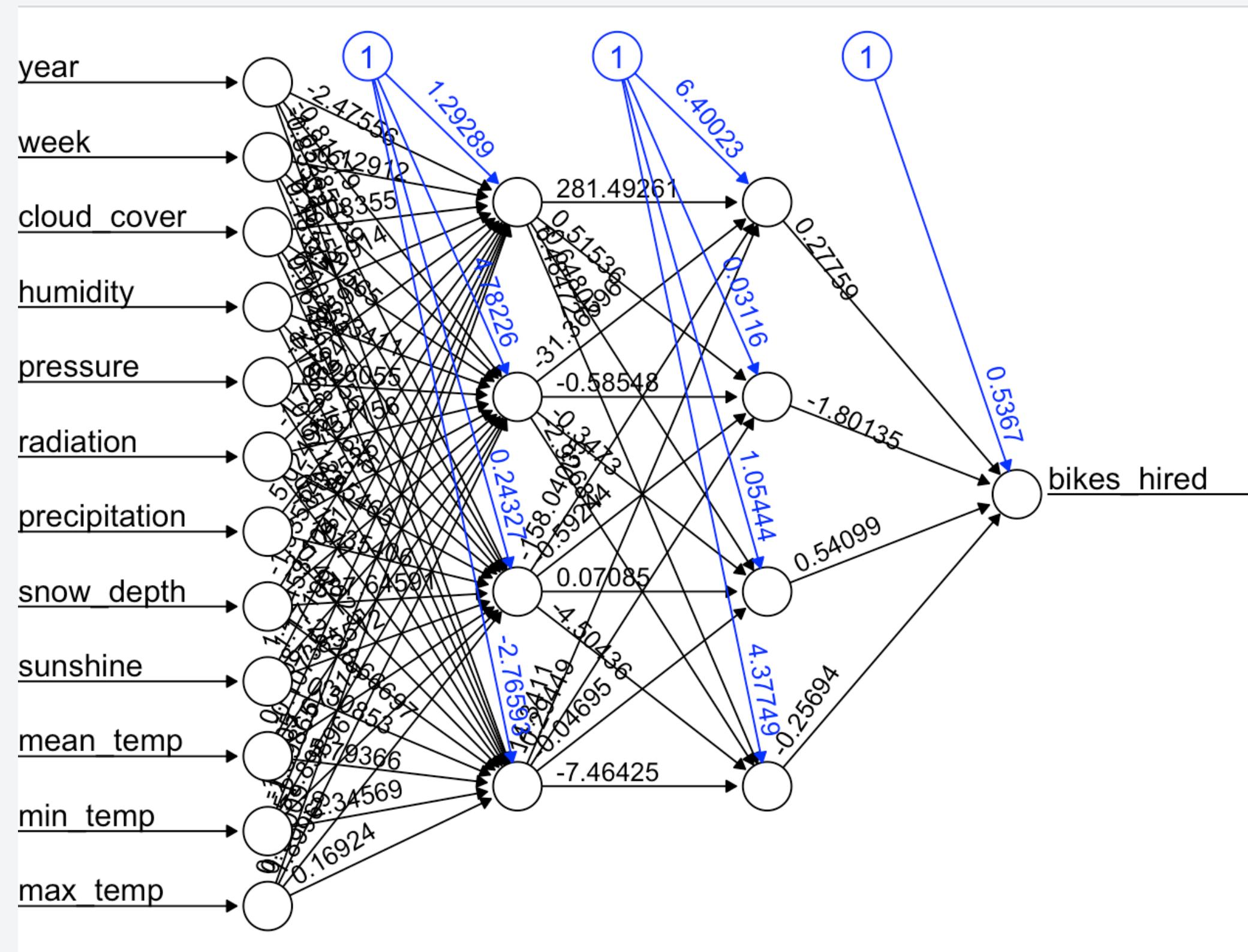
- Hidden layers: 2
- Neurons: 4
- 71.86% of the bike rentals can be explained by these variables
- Higher error metrics than linear regression; thus this model does not perform as well

	Training	Testing
R2	0.7186374	0.1709133
RMSE	0.07022687	0.2296976
MAPE	Inf	0.4568433
MAE	0.05576964	0.18582
MSE	0.004931813	0.05276098

* This is a black box model; thus the only error metrics are relevant

** Higher complexity in network achieves, better error metrics but is more computationally expensive

APPENDIX: NEURAL NETWORK 1



APPENDIX: NEURAL NETWORK 2

Model 2: Dropped Insignificant Features

Remove insignificant variables found in linear regression to try improve neural network

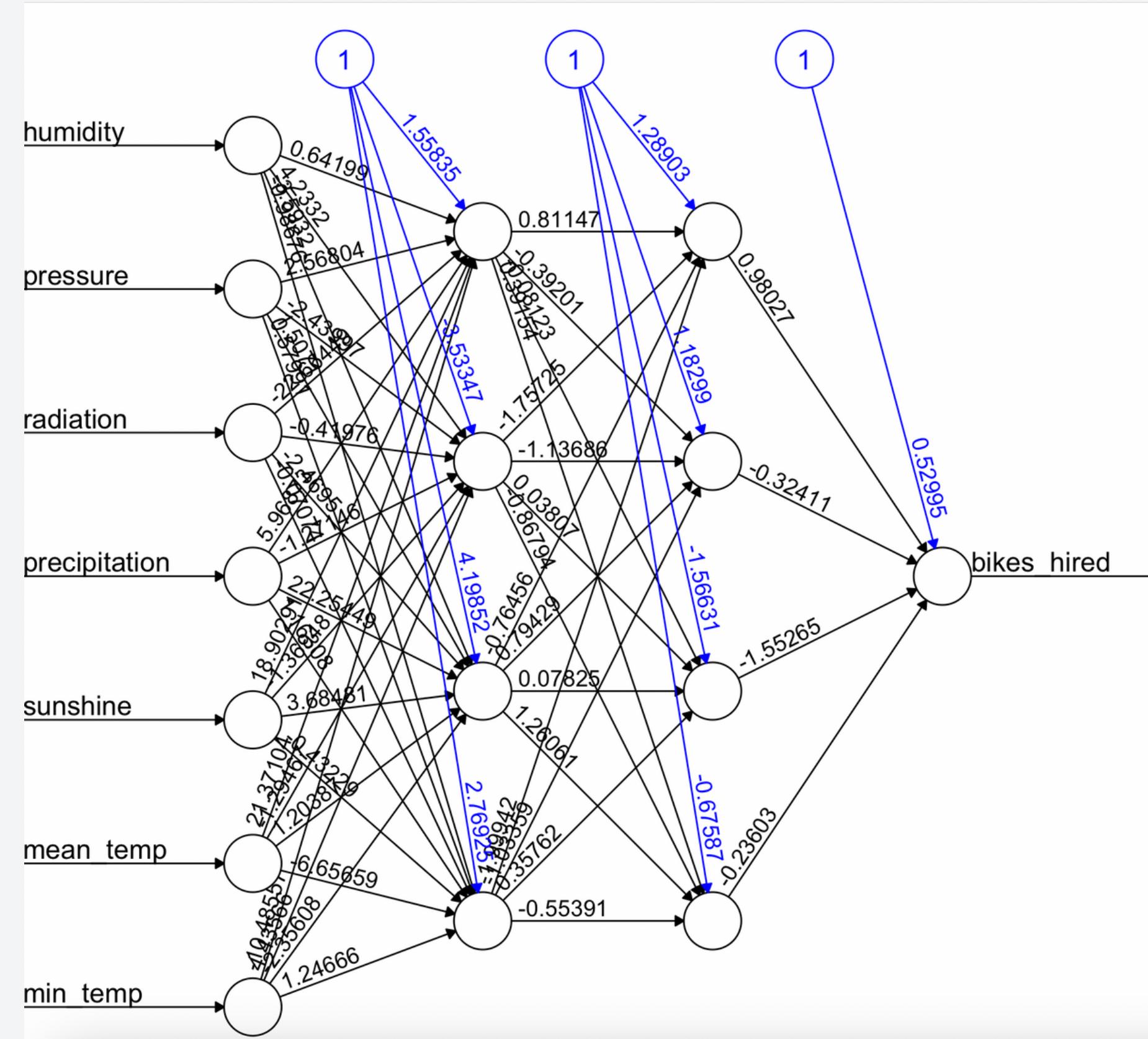
- Hidden layers: 2
- Neurons: 4
- Removed variables:
 - Snow depth
 - Cloud cover
 - Radiation
 - Week
- 66.13% of the bike rentals can be explained by these variables
- High error metrics than previous model

	Training	Testing
R2	0.6594119	0.6086737
RMSE	0.07726535	0.127202
MAPE	Inf	0.2654279
MAE	0.06228418	0.1027758
MSE	0.005969935	0.01618034

* This is a black box model; thus the only error metrics are relevant

** Higher complexity in network achieves, better error metrics but is more computationally expensive

APPENDIX: NEURAL NETWORK 2



APPENDIX: MODELS AND FINDINGS FROM RANDOM SAMPLE

CREATE MODELS WITH RANDOM SAMPLING TO ACCOUNT FOR EFFECT OF DAY OF THE WEEK AND MONTH OF THE YEAR

Model	Training R2	Testing R2	Insights
1. Linear Regression (all variables**)	73.59%	73.43%	Most significant predictor variables (Date & Weather)
2. Linear Regression (only significant variables***)	73.55%	73.41%	Demand prediction (best model to predict bike demand)

Best model to predict how many bikes are hired per day is created using **random sampling** and accounts for **month of the year, day of the week and weather patterns**.

APPENDIX: LOGISTIC REGRESSION: RANDOM SAMPLING AND ALL VARIABLES

- Random sample to account for seasonality in bike hires
- Randomly sample: 80:20
- Good error metrics across training and testing

	Training	Testing
R2	0.7359478	0.7343214
RMSE	5054.528	5108.53
MAPE	0.1818826	0.1827067
MAE	3732.39	3763.723
MSE	25548257	26097080

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.385e+06	5.740e+04	-24.135	< 2e-16 ***
year	6.617e+02	2.820e+01	23.465	< 2e-16 ***
wdayMon	-8.247e+02	3.342e+02	-2.467	0.01366 *
wdaySat	-2.466e+03	3.321e+02	-7.425	1.43e-13 ***
wdaySun	-4.765e+03	3.376e+02	-14.115	< 2e-16 ***
wdayThu	1.060e+03	3.342e+02	3.173	0.00152 **
wdayTue	9.211e+02	3.321e+02	2.774	0.00558 **
wdayWed	9.880e+02	3.299e+02	2.995	0.00277 **
monthAug	3.386e+03	6.689e+02	5.062	4.38e-07 ***
monthDec	-2.087e+03	9.922e+02	-2.103	0.03555 *
monthFeb	-1.383e+03	5.625e+02	-2.459	0.01398 *
monthJan	-1.420e+03	6.158e+02	-2.306	0.02119 *
monthJul	4.381e+03	6.139e+02	7.136	1.18e-12 ***
monthJun	4.569e+03	5.443e+02	8.395	< 2e-16 ***
monthMar	-6.701e+02	4.684e+02	-1.430	0.15269
monthMay	3.034e+03	4.682e+02	6.481	1.05e-10 ***
monthNov	2.224e+03	9.262e+02	2.401	0.01641 *
monthOct	3.725e+03	8.220e+02	4.532	6.06e-06 ***
monthSep	4.072e+03	7.318e+02	5.565	2.84e-08 ***
week	-1.162e+01	2.433e+01	-0.477	0.63311
cloud_cover	-1.500e+01	7.418e+01	-0.202	0.83973
humidity	-8.143e+01	1.415e+01	-5.756	9.40e-09 ***
pressure	7.269e+00	9.345e-01	7.778	9.81e-15 ***
radiation	-1.054e+01	3.534e+00	-2.984	0.00287 **
precipitation	-3.549e+01	2.803e+00	-12.661	< 2e-16 ***
snow_depth	-6.654e+02	3.665e+02	-1.816	0.06952 .
sunshine	5.052e+01	7.018e+00	7.198	7.55e-13 ***
mean_temp	1.370e+03	9.688e+01	14.142	< 2e-16 ***
min_temp	-6.887e+02	8.054e+01	-8.551	< 2e-16 ***
max_temp	-5.828e+01	3.753e+01	-1.553	0.12062

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5078 on 3243 degrees of freedom
Multiple R-squared: 0.7359, Adjusted R-squared: 0.7336
F-statistic: 311.7 on 29 and 3243 DF, p-value: < 2.2e-16

APPENDIX: LOGISTIC REGRESSION: RANDOM SAMPLING AND ONLY SIGNIFICANT VARIABLES

- Removed variables:
 - Snow depth
 - Cloud cover
 - Week
- Good error metrics

	Training	Testing
R2	0.7354659	0.7340808
RMSE	5059.138	5110.844
MAPE	0.1817842	0.1824894
MAE	3735.101	3763.569
MSE	25594882	26120729

Residuals:

	Min	1Q	Median	3Q	Max
	-24427	-2786	261	2888	31569

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.386e+06	5.693e+04	-24.343	< 2e-16 ***
year	6.624e+02	2.798e+01	23.671	< 2e-16 ***
wdayMon	-8.469e+02	3.340e+02	-2.536	0.011272 *
wdaySat	-2.469e+03	3.320e+02	-7.435	1.33e-13 ***
wdaySun	-4.792e+03	3.375e+02	-14.198	< 2e-16 ***
wdayThu	1.062e+03	3.342e+02	3.178	0.001495 **
wdayTue	9.139e+02	3.319e+02	2.754	0.005924 **
wdayWed	9.857e+02	3.298e+02	2.989	0.002821 **
monthAug	2.988e+03	5.165e+02	5.786	7.89e-09 ***
monthDec	-2.394e+03	5.459e+02	-4.385	1.20e-05 ***
monthFeb	-1.207e+03	5.193e+02	-2.324	0.020210 *
monthJan	-1.149e+03	5.470e+02	-2.101	0.035693 *
monthJul	4.005e+03	5.220e+02	7.672	2.23e-14 ***
monthJun	4.306e+03	5.002e+02	8.609	< 2e-16 ***
monthMar	-5.493e+02	4.558e+02	-1.205	0.228267
monthMay	2.895e+03	4.554e+02	6.358	2.33e-10 ***
monthNov	1.969e+03	5.373e+02	3.665	0.000251 ***
monthOct	3.400e+03	5.150e+02	6.602	4.71e-11 ***
monthSep	3.703e+03	5.045e+02	7.341	2.67e-13 ***
humidity	-8.235e+01	1.411e+01	-5.834	5.93e-09 ***
pressure	7.115e+00	9.243e-01	7.698	1.82e-14 ***
precipitation	-3.518e+01	2.799e+00	-12.570	< 2e-16 ***
radiation	-1.083e+01	3.493e+00	-3.102	0.001940 **
sunshine	5.187e+01	5.560e+00	9.329	< 2e-16 ***
mean_temp	1.292e+03	7.902e+01	16.345	< 2e-16 ***
min_temp	-6.445e+02	7.414e+01	-8.694	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Residual standard error: 5079 on 3247 degrees of freedom				
Multiple R-squared: 0.7355, Adjusted R-squared: 0.7334				
F-statistic: 361.1 on 25 and 3247 DF, p-value: < 2.2e-16				

APPENDIX: GITHUB LINK

https://github.com/KrishanTCD/TCDBA_Operations_Group_4