

Intro 1

People often rely on machine learning model outputs to make decisions.

Many factors can contribute to a machine learning model's output. For example, the output of a rain-predicting model can rely on factors such as the current temperature and wind speed.

Computer scientists refer to these factors as **model explanations**.

We will teach you how to interpret these explanations and ask you questions about them.

Intro 2

Someone designed a machine learning model to predict whether it is a good idea to put on a coat or not.

It calculates the probability that you should put on a coat using

the current temperature, wind speed, and precipitation.

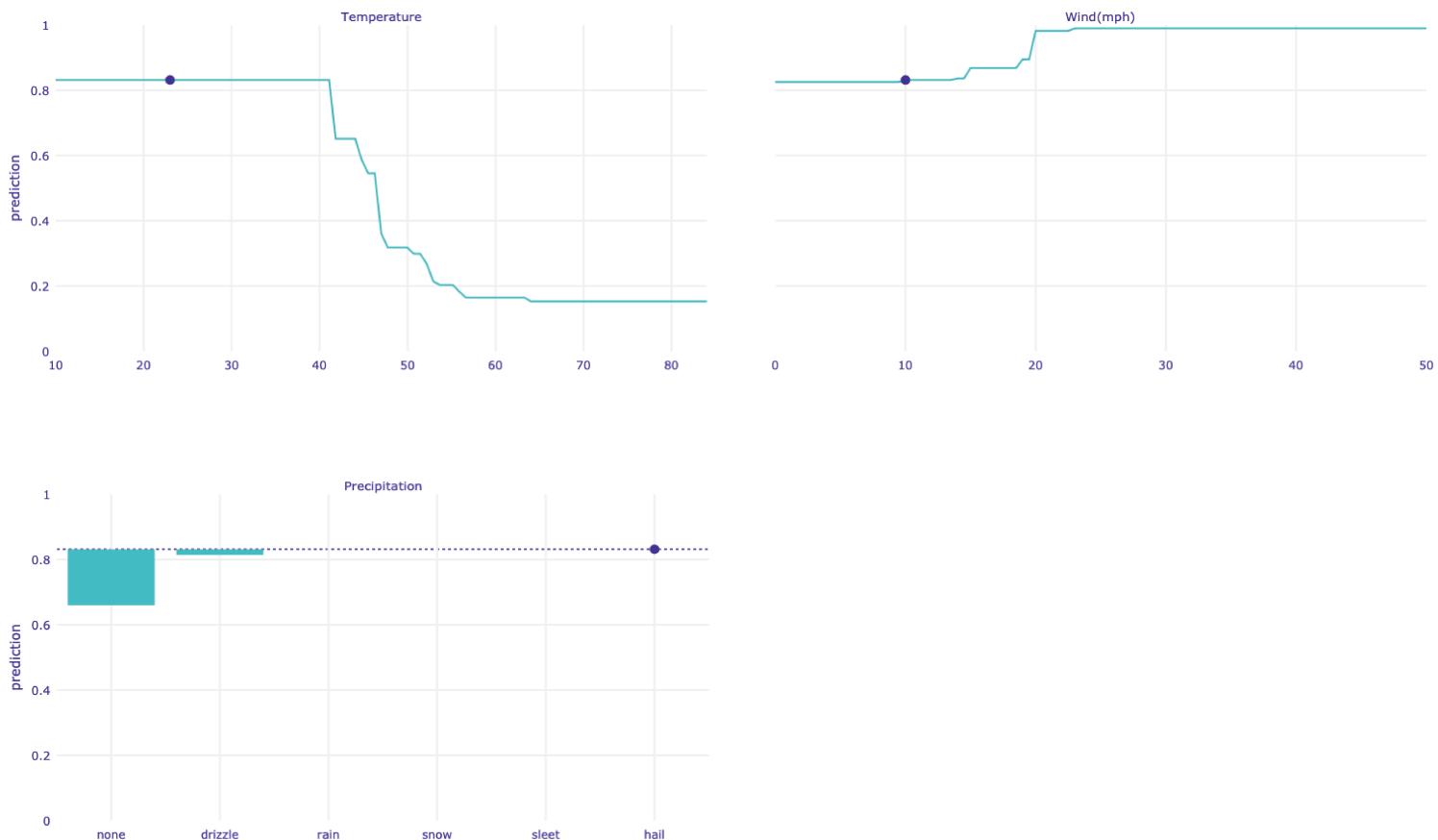
If that probability is greater than or equal to 0.5, then the model will recommend that you put on a coat. If the probability is less than 0.5, then the model will recommend that you do NOT put on a coat.

Intro 3

Below, you can see a visual explanation for one instance of the model prediction, based on some input values for the three factors the model considers (temperature, wind speed, and precipitation).

Each chart in this explanation shows you how the model's prediction will change if **only** that feature is changed.

Model Prediction: 0.861



Let's take a closer look at this visual explanation.

Intro 4

You can see two types of charts below. The first two are line

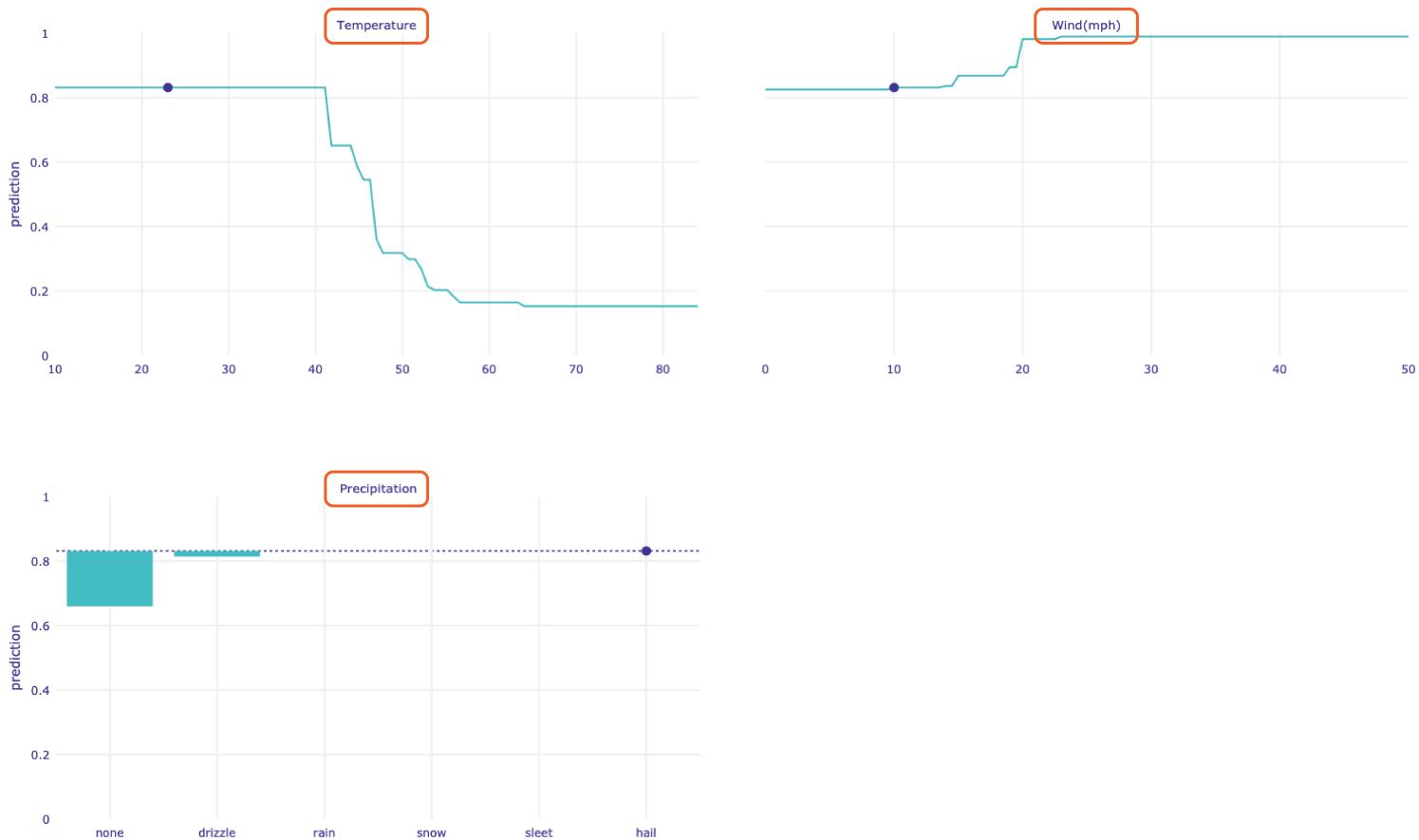
charts, and the second is a bar chart.

Each chart has the name of the feature it corresponds to listed above.

Line charts correspond to **numerical** features – features that can take on a range of numerical values.

Bar charts correspond to **categorical** features – features which have values that are not numbers.

Model Prediction: 0.861

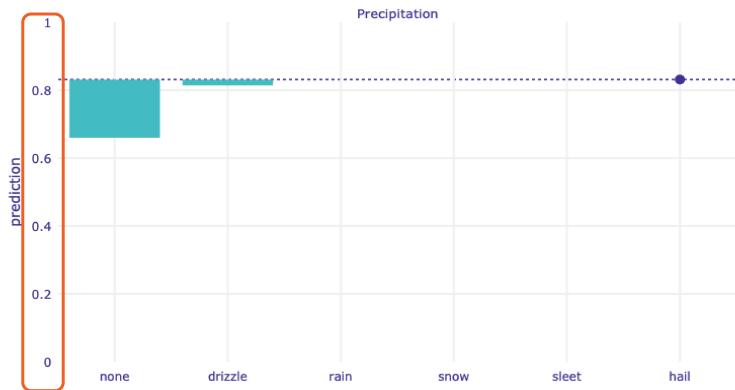
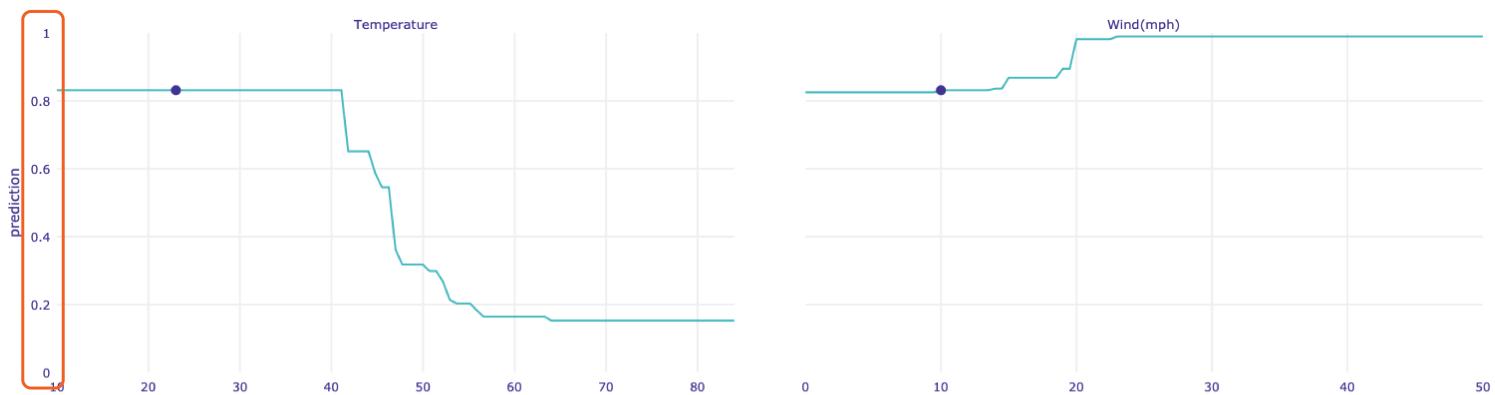


Intro 5

The vertical axis on both the line and the bar charts shows the probability value predicted by the model.

This probability describes whether it is a good idea to put on a coat or not (probability ≥ 0.5 , good idea to put on a coat; probability < 0.5 , NOT a good idea).

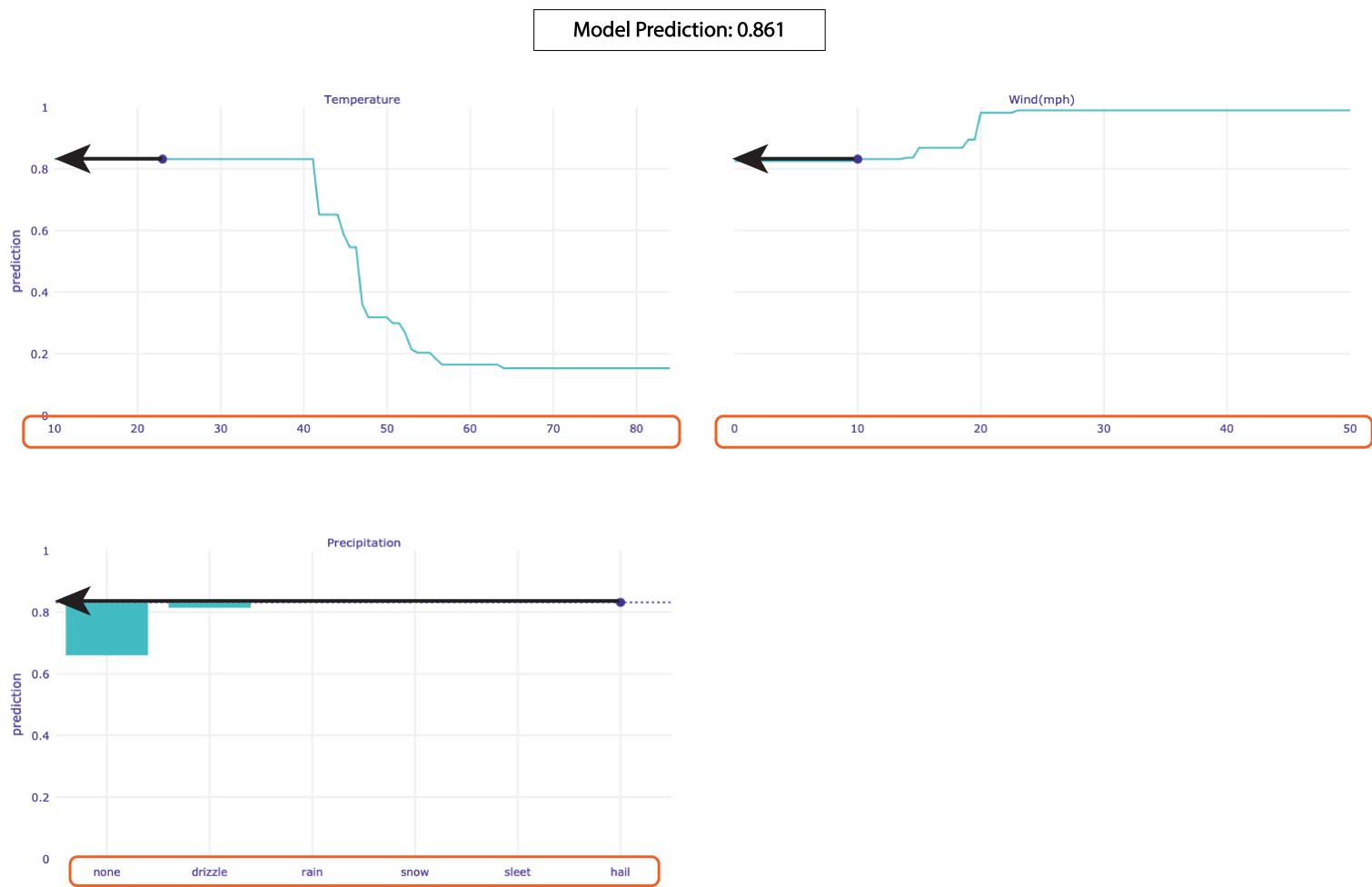
Model Prediction: 0.861



Intro 6

The horizontal axis on both the line and the bar charts shows the

possible values for each feature. For example, the x axis for temperature ranges from 10 to 80, because that is the range of values that temperature can take for this model. Since precipitation is categorical, it has six possible values – none, drizzle, rain, snow, sleet, and hail.



Intro 7

You can put different values of temperature, wind, and precipitation into your model to generate a model prediction.

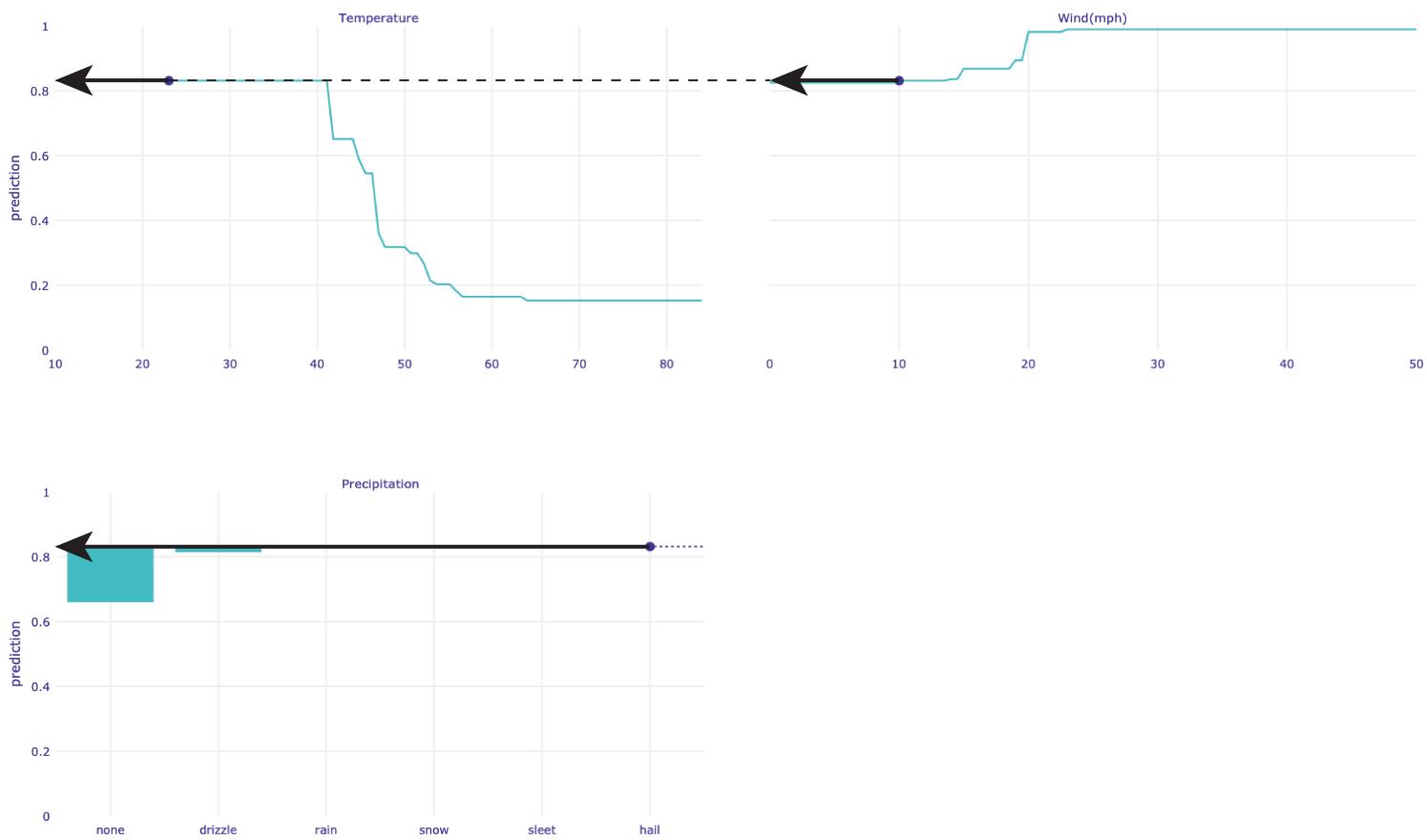
The position of blue dot in each chart shows the the value of each feature which was put in to generate the **current prediction** (x-axis), and value of the prediction itself (y-axis).

The horizontal position of the dot is the value of the feature. The vertical position is the resulting model prediction – you can look **to the left of the dot** to see what the model predicted.

You are seeing a description for a **single model prediction**, so the y-axis prediction value where the blue dot is will be **the same** for every feature.

The exact value of this prediction is in the Model Prediction box at the top of the visualization. Here, that value is 0.861.

Model Prediction: 0.861



Intro 7.5

You can put different values of temperature, wind, and precipitation into your model to generate a model prediction.

The position of blue dot in each chart shows the the value of each feature which was put in to generate the **current prediction** (x-axis), and value of the prediction itself (y-axis).

The horizontal position of the dot is the value of the feature. You can look **below the dot** to see what the value of temperature, wind or precipitation is.

You can mouse over a chart to see a small table with a "Value" column. This is the exact value of the feature. Try it!

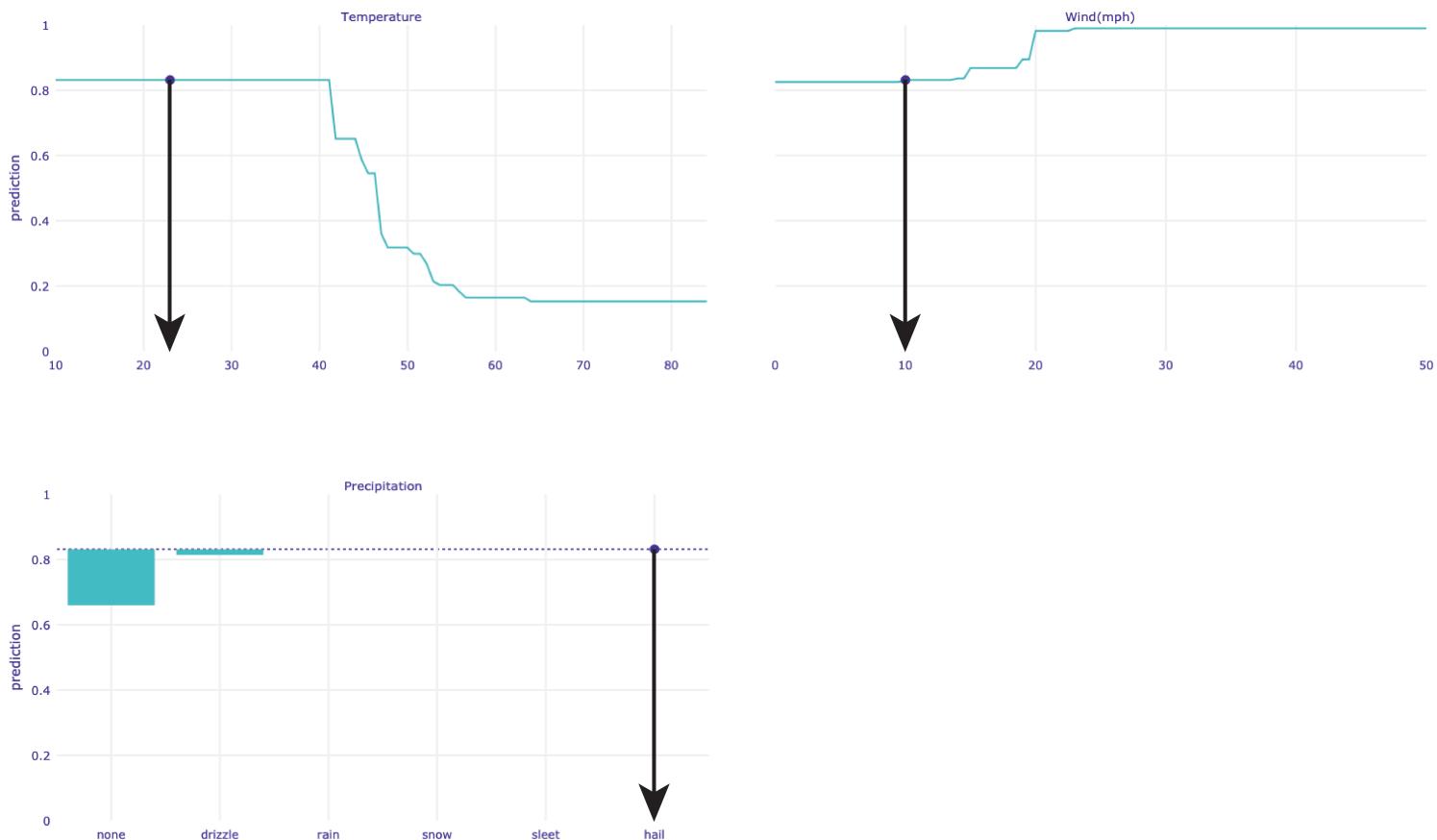
Here, our feature values are:

Temperature = 23

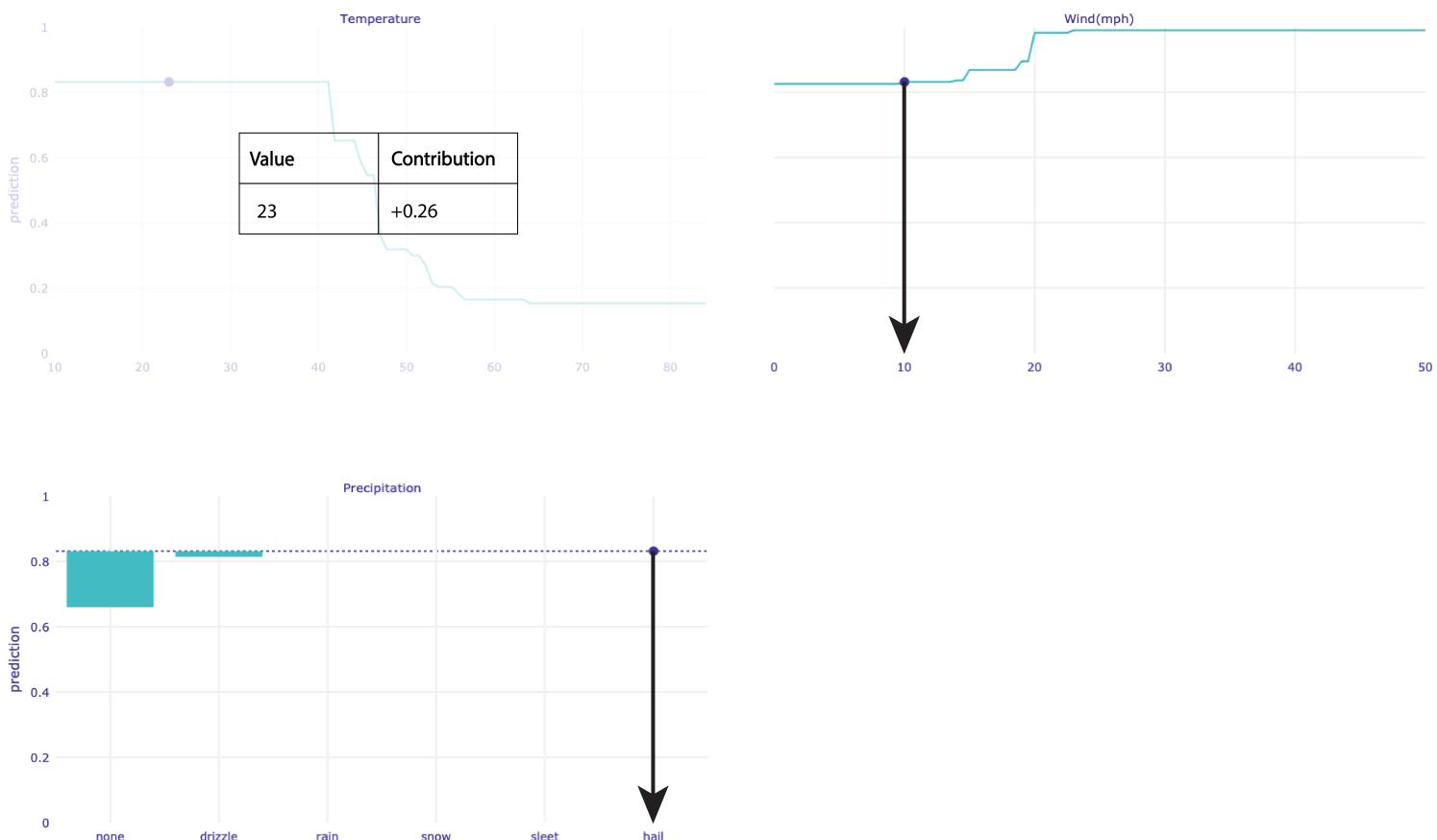
Wind(mph) = 10

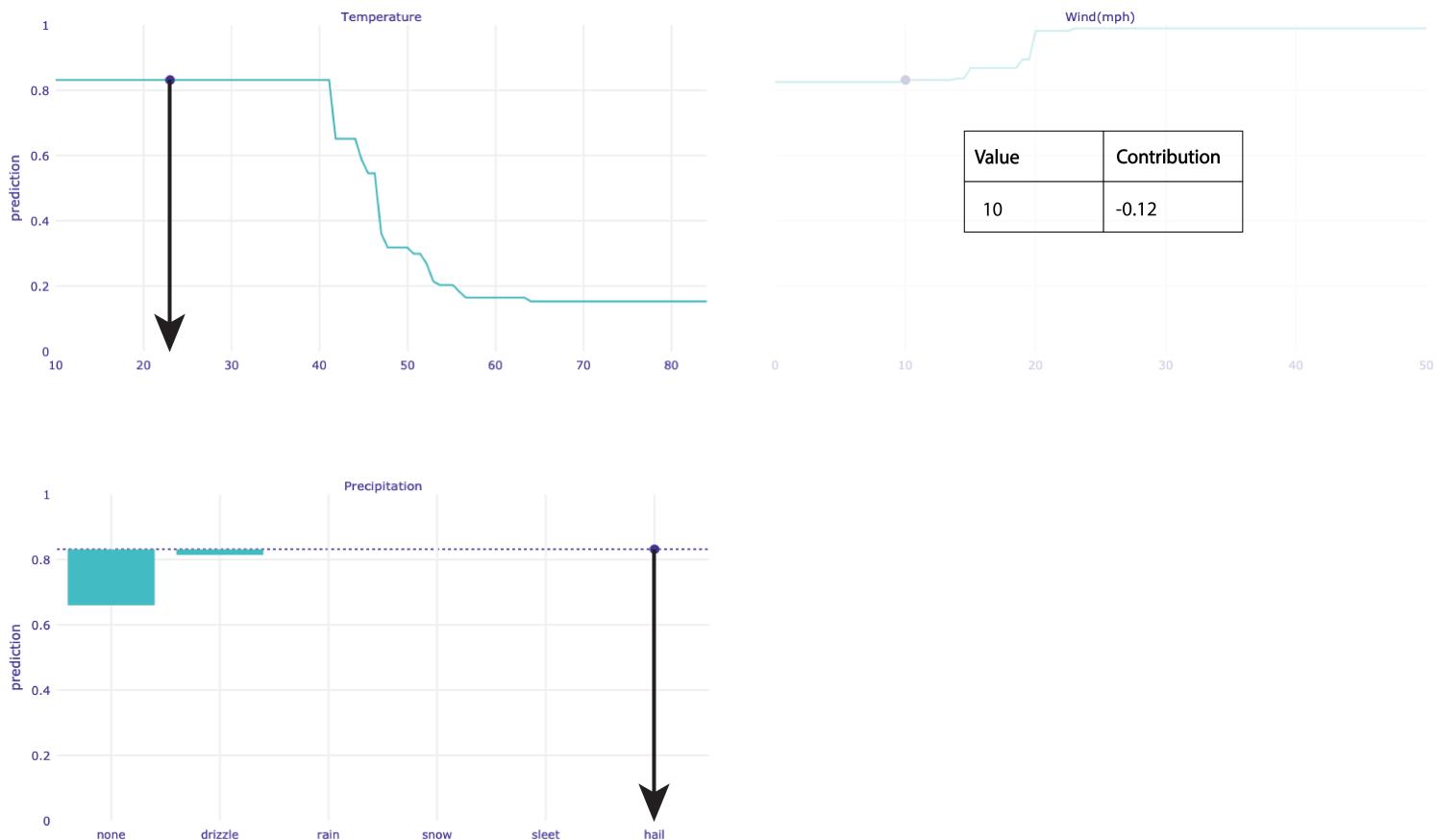
Precipitation = hail

Model Prediction: 0.861

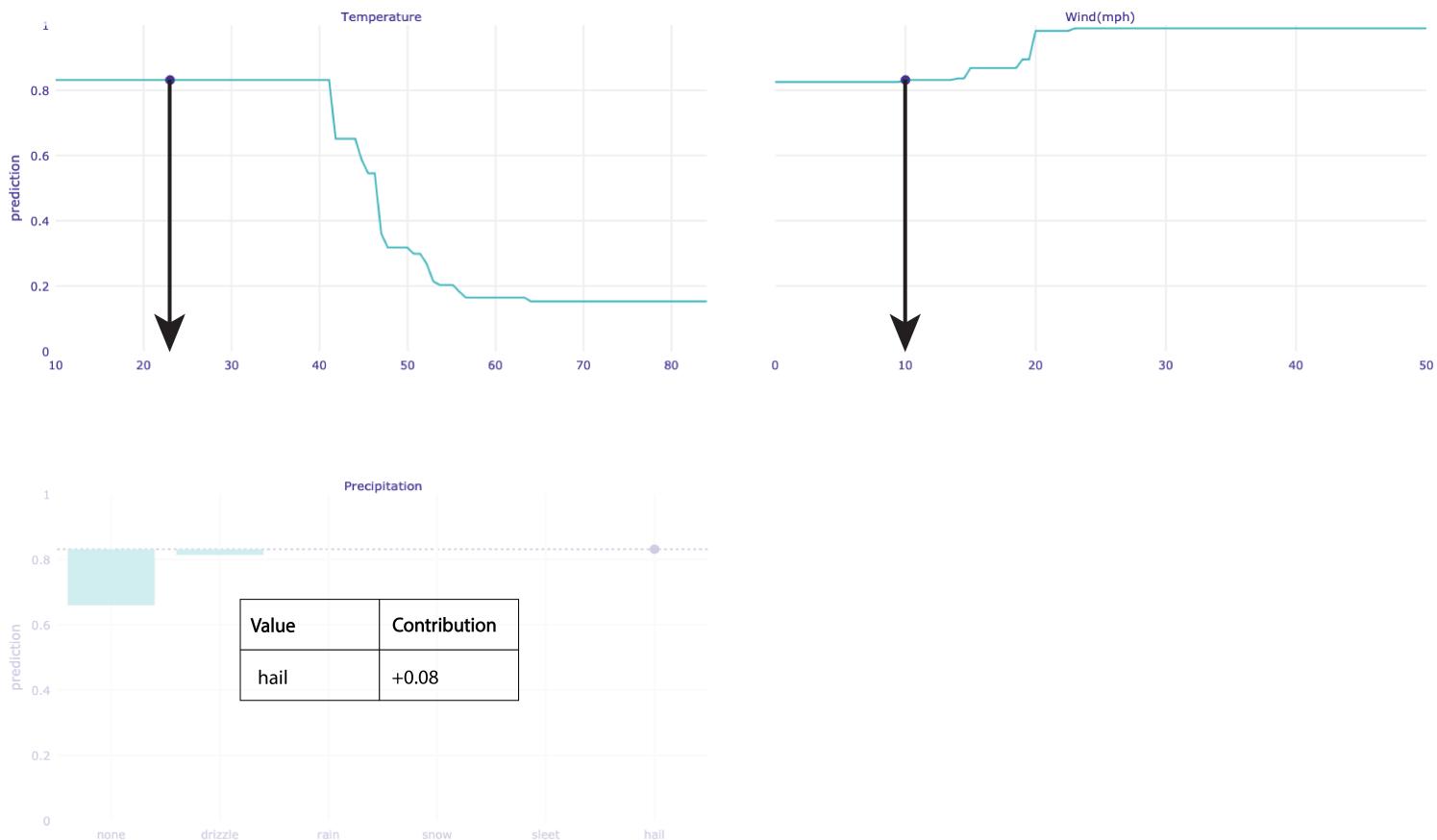


Model Prediction: 0.861



Model Prediction: 0.861

Model Prediction: 0.861

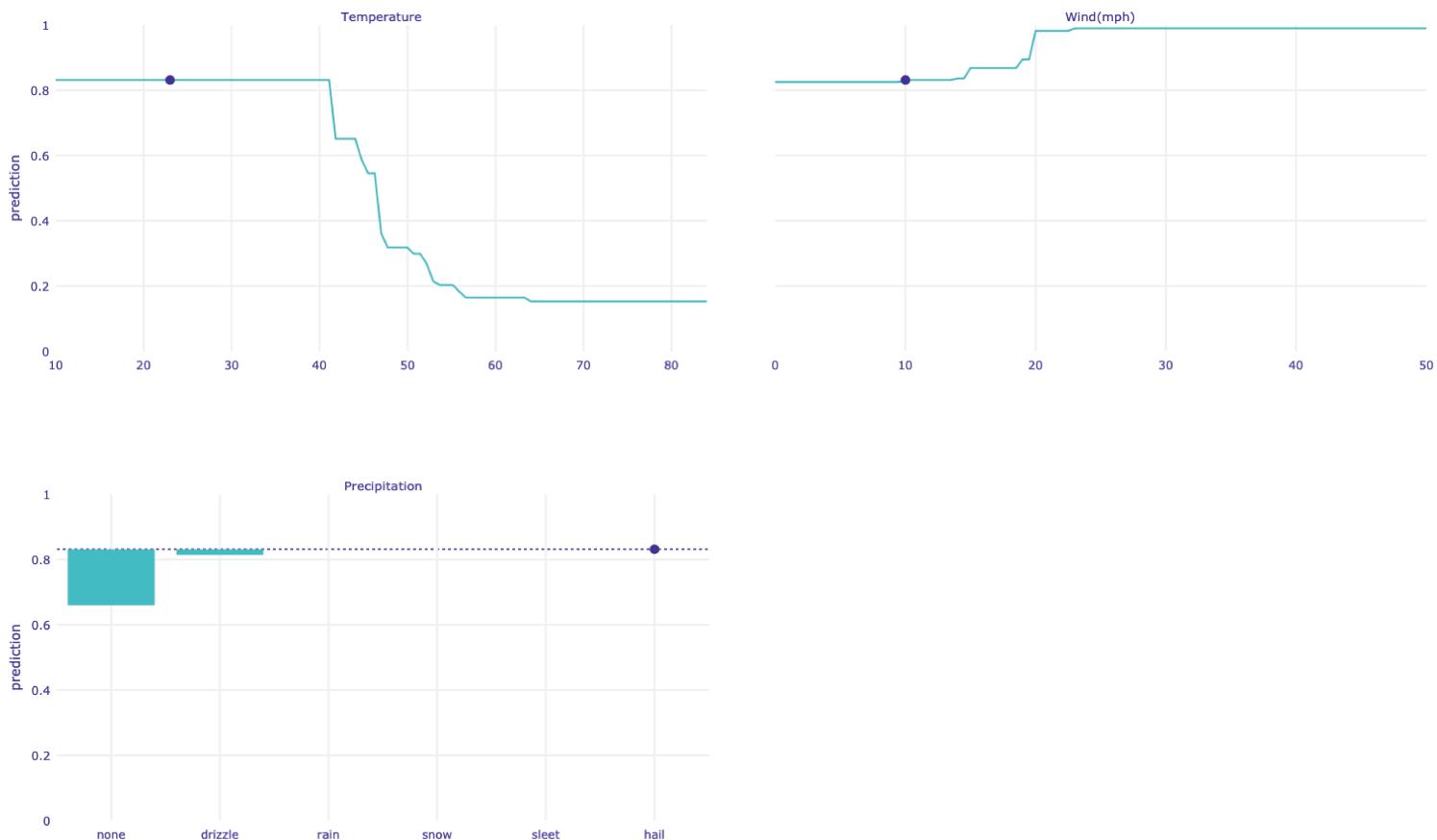


Intro Test 1

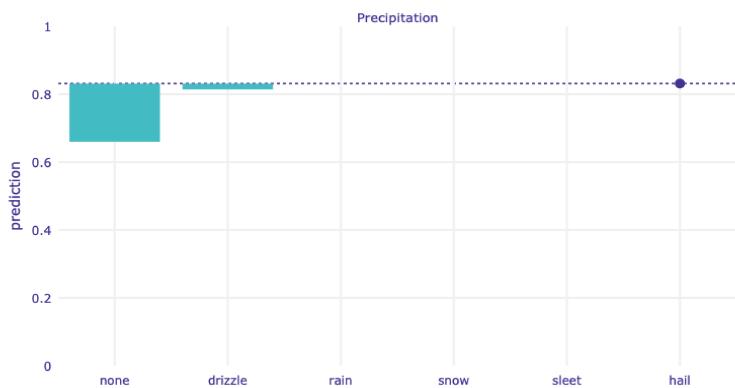
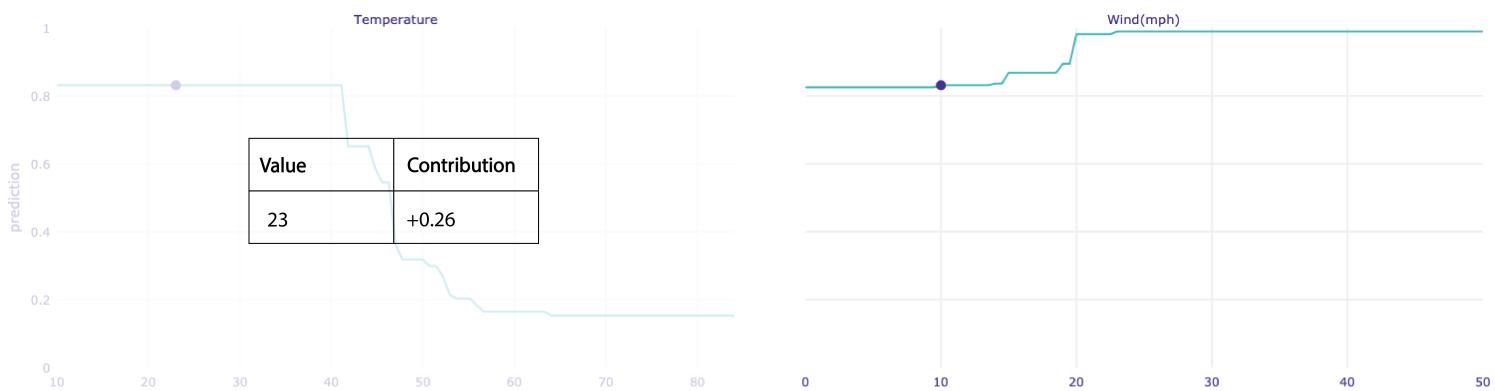
In the example below, what will the model predict?

- YES, you should wear a coat
- NO, do not wear a coat

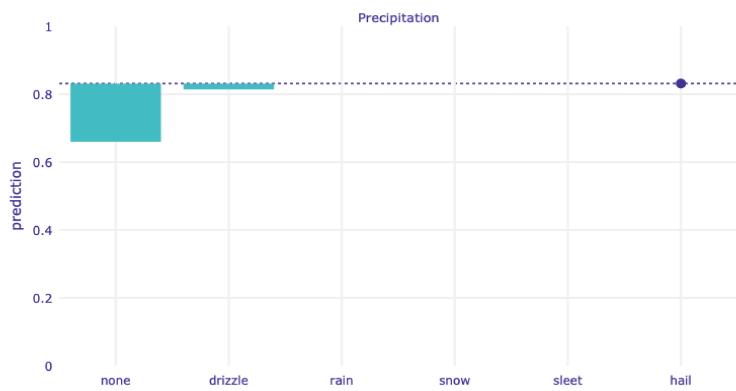
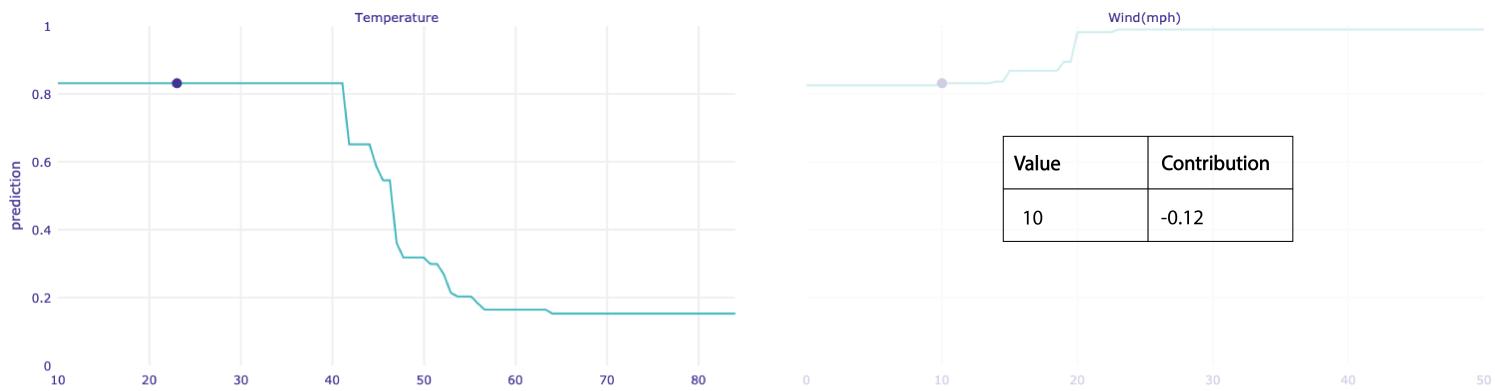
Model Prediction: 0.861

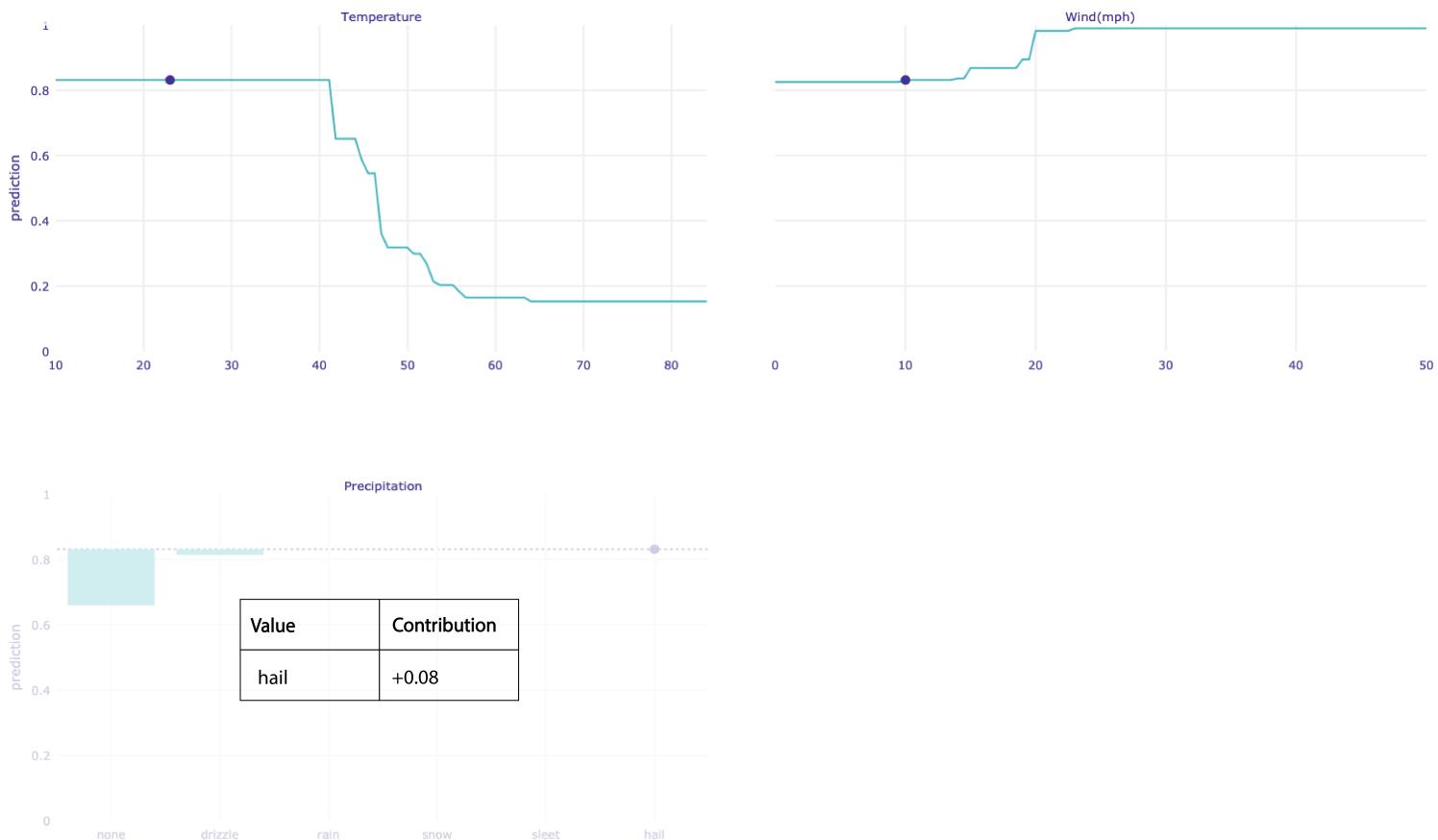


Model Prediction: 0.861



Model Prediction: 0.861

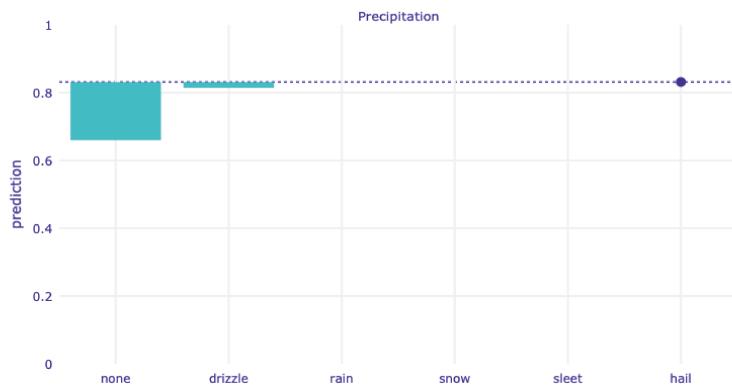
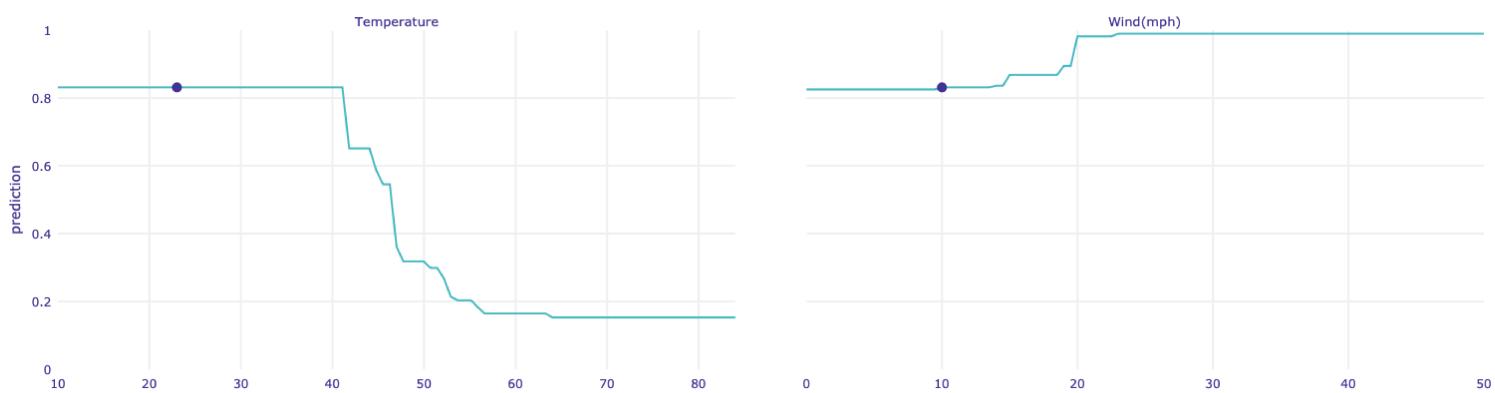


Model Prediction: 0.861

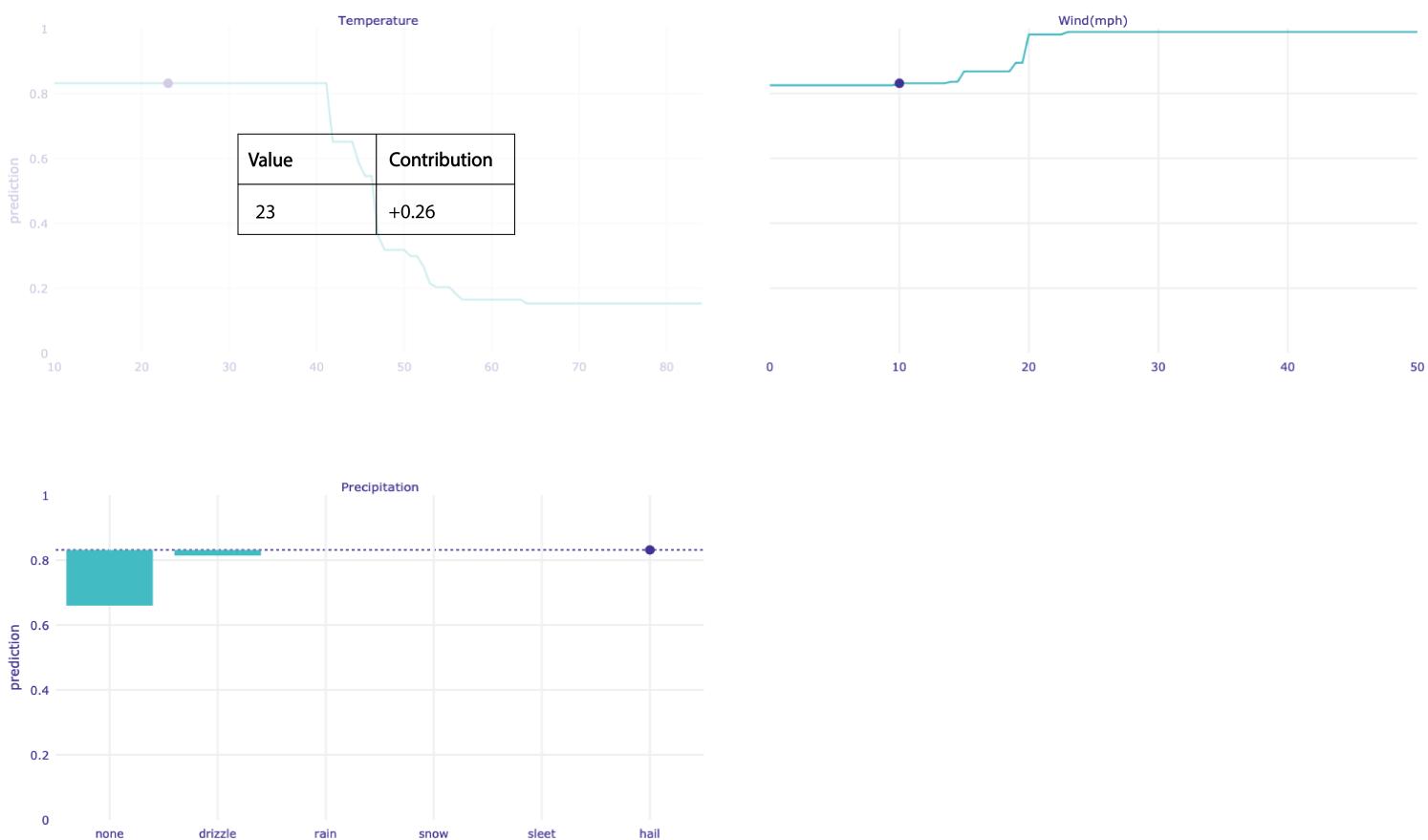
Correct. In this case, the model prediction is 0.861, which is larger than 0.5, so the model will return YES.

Not quite. In this case, the model prediction is 0.861, which is larger than 0.5, so the model will return YES.

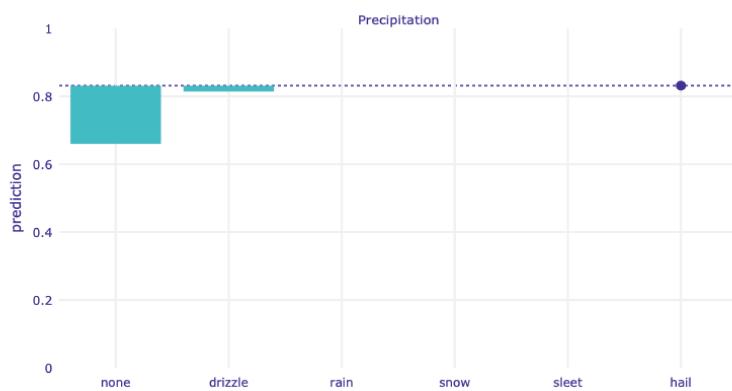
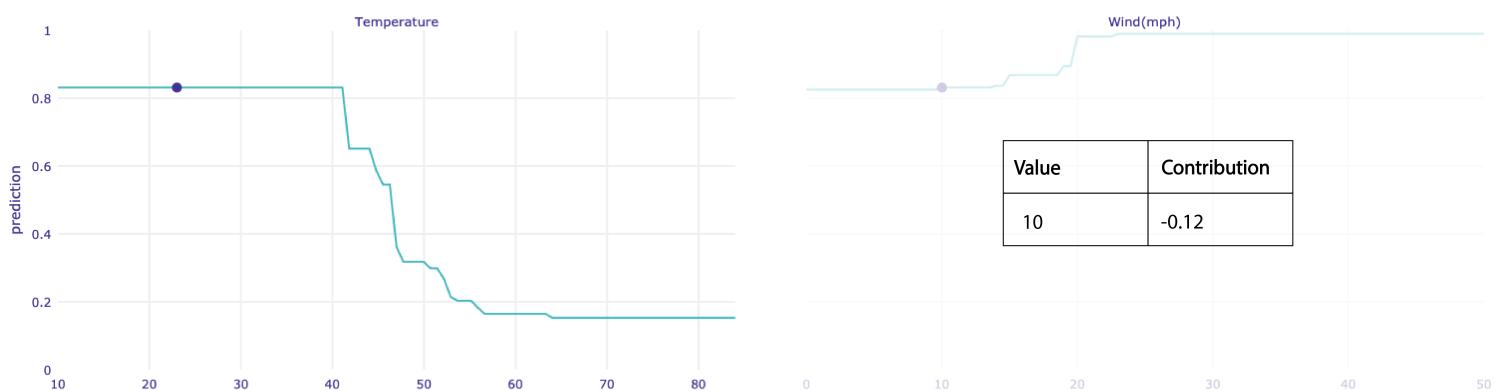
Model Prediction: 0.861



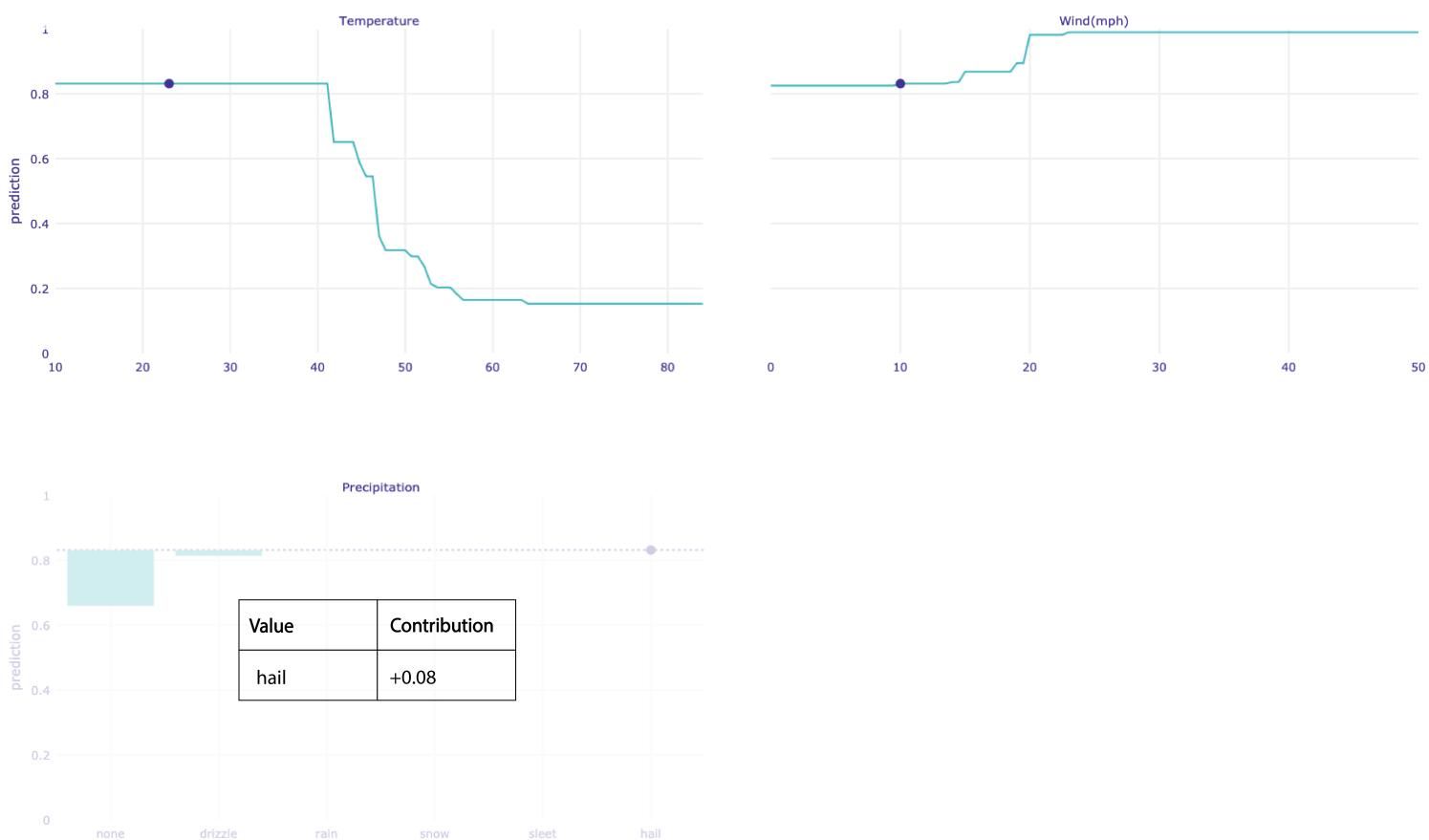
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861



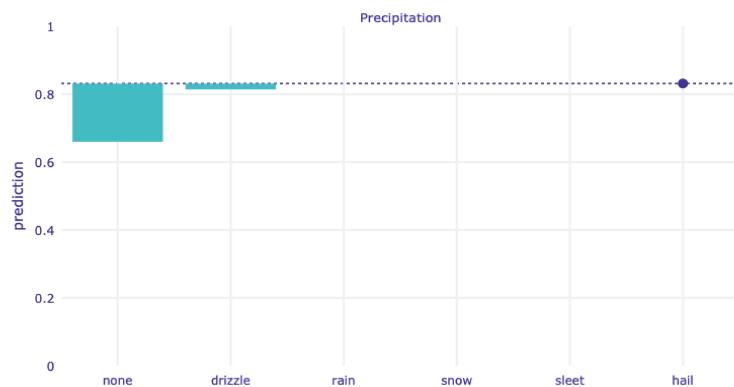
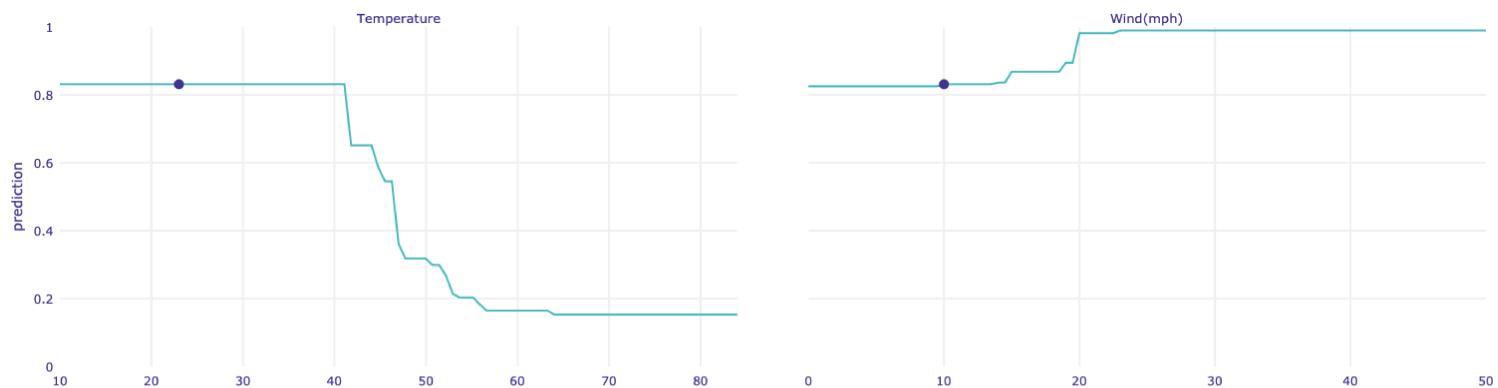
Intro Test 2

As another review, by looking at the explanation image, please select the value for **precipitation** input into the model:

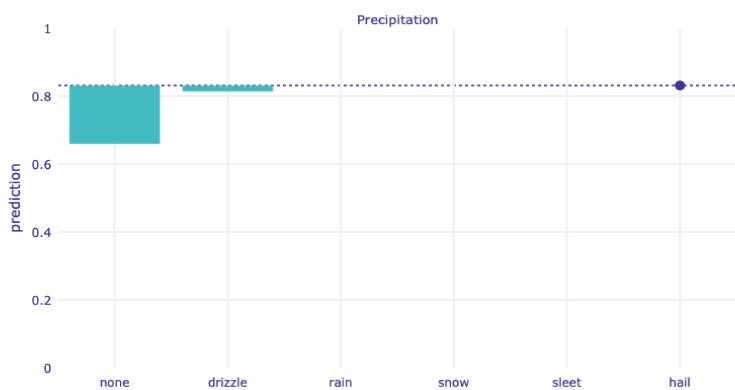
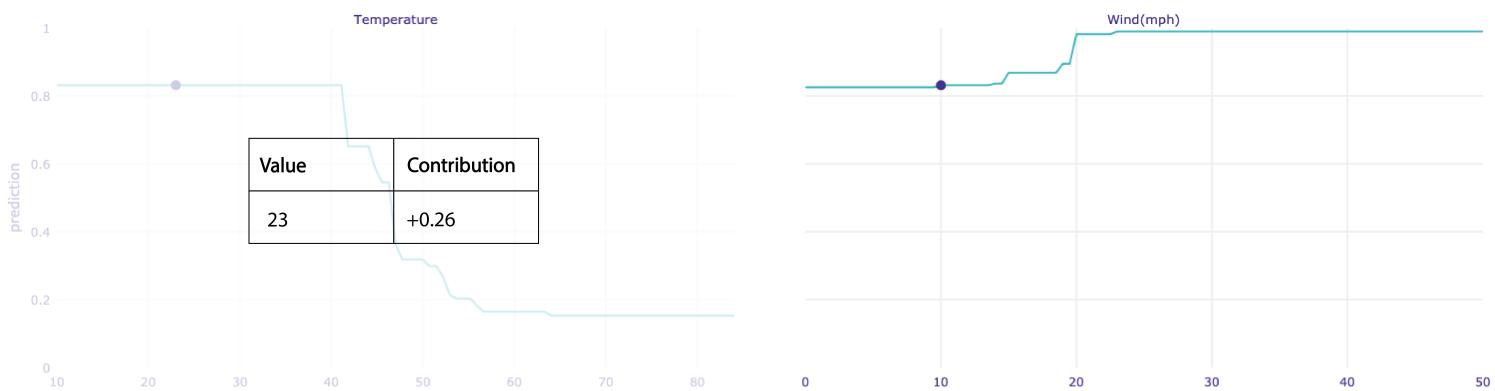
- sleet
- snow
- hail

- rain
- none

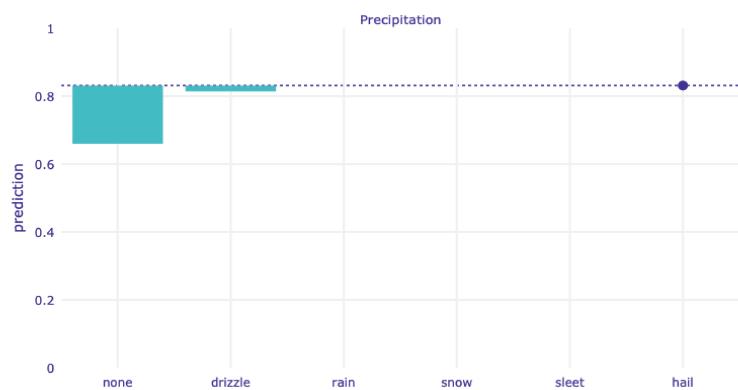
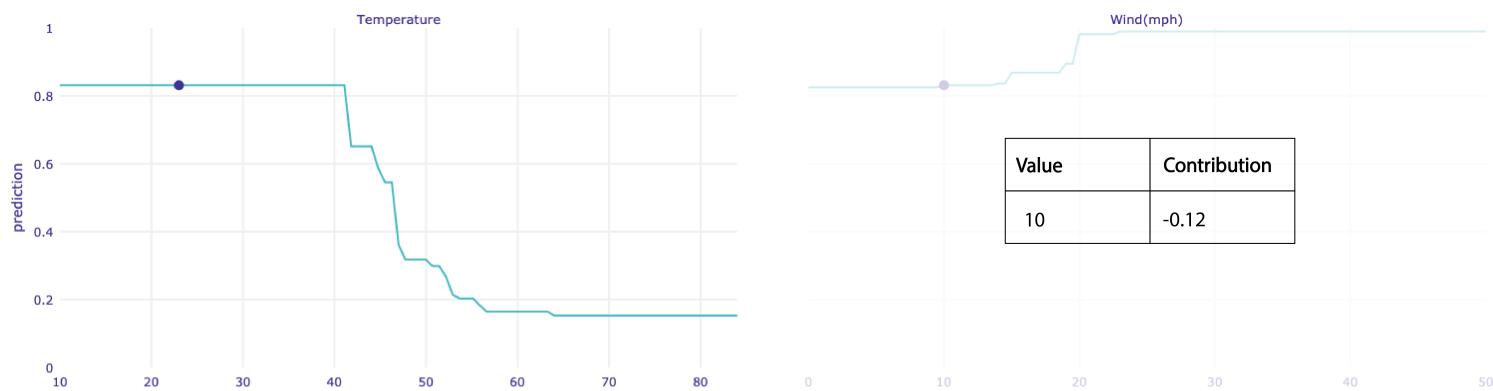
Model Prediction: 0.861



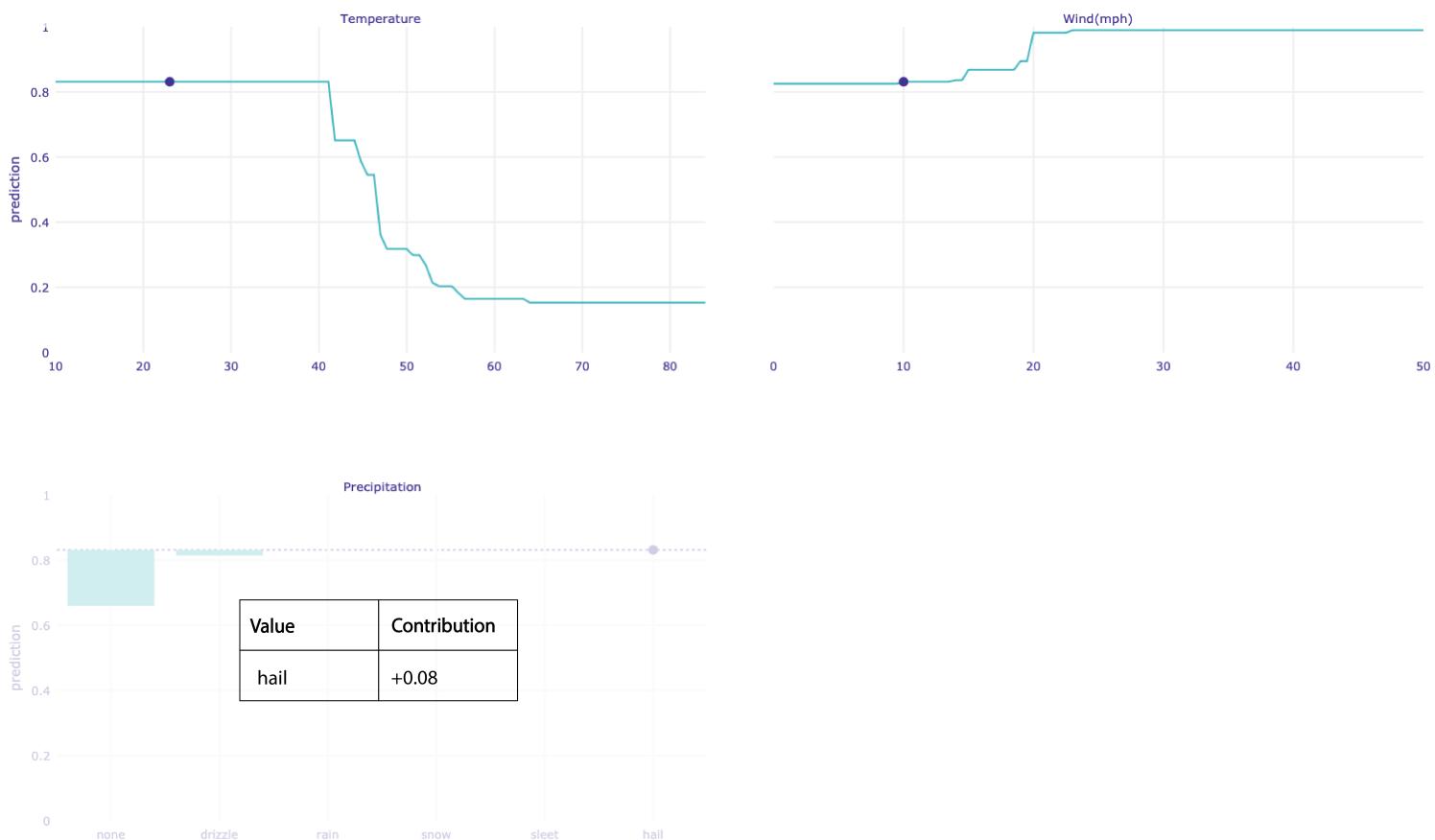
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861

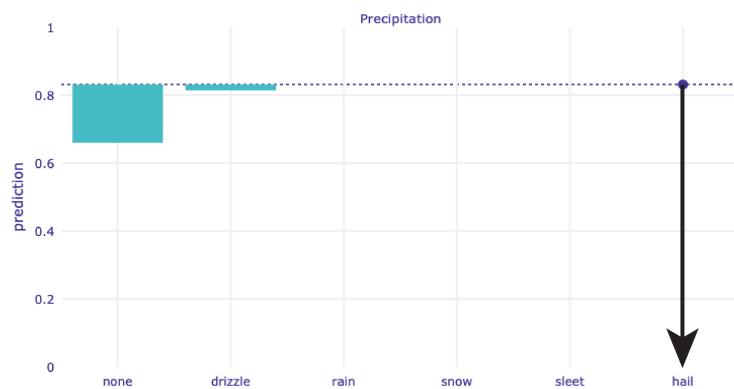
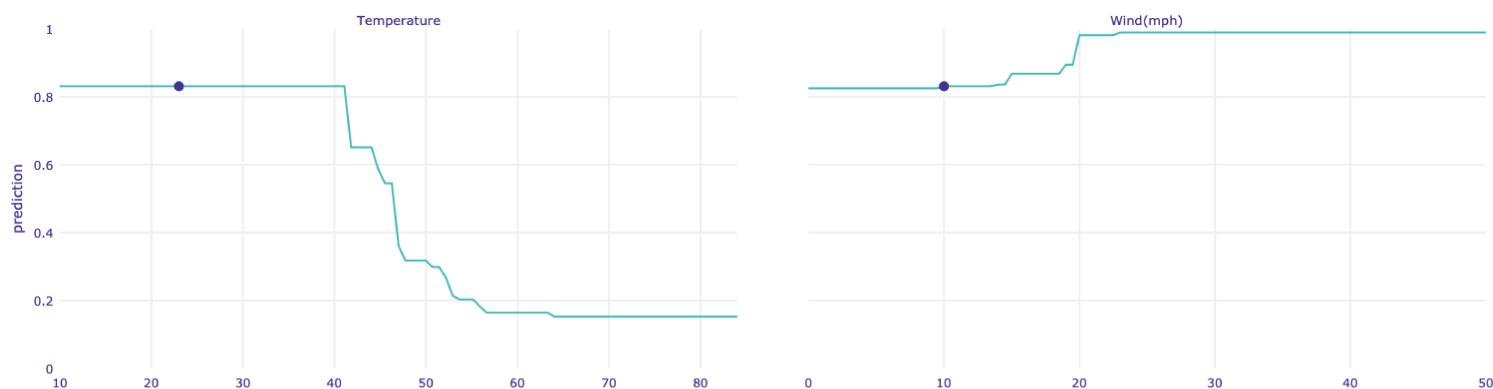


Correct – the value is in the table when you mouse over the precipitation chart, and below the blue dot in the precipitation chart. This value is **hail**.

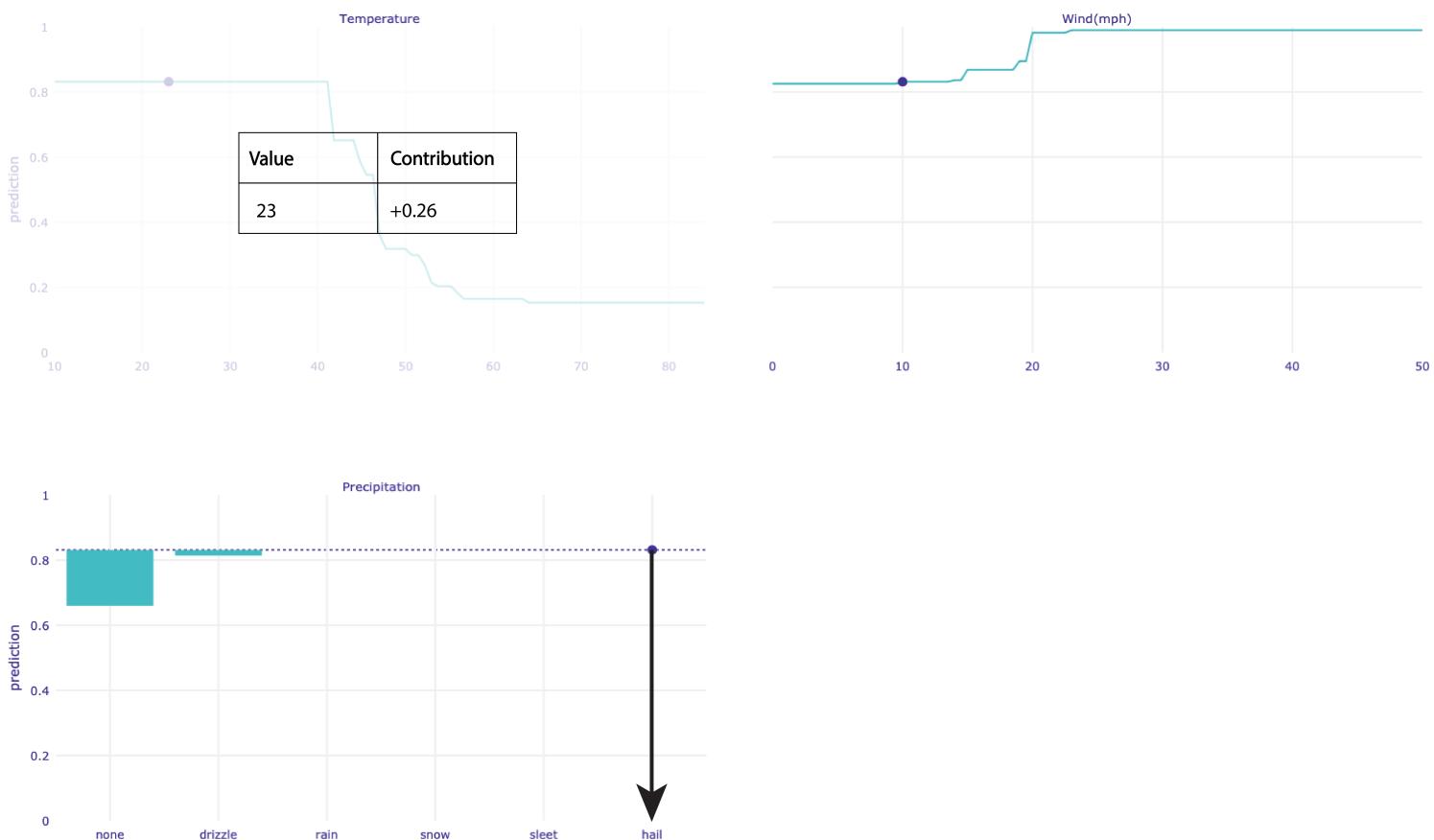
Not quite – the value is in the table when you mouse over the

precipitation chart, and below the blue dot in the precipitation chart. This value is hail.

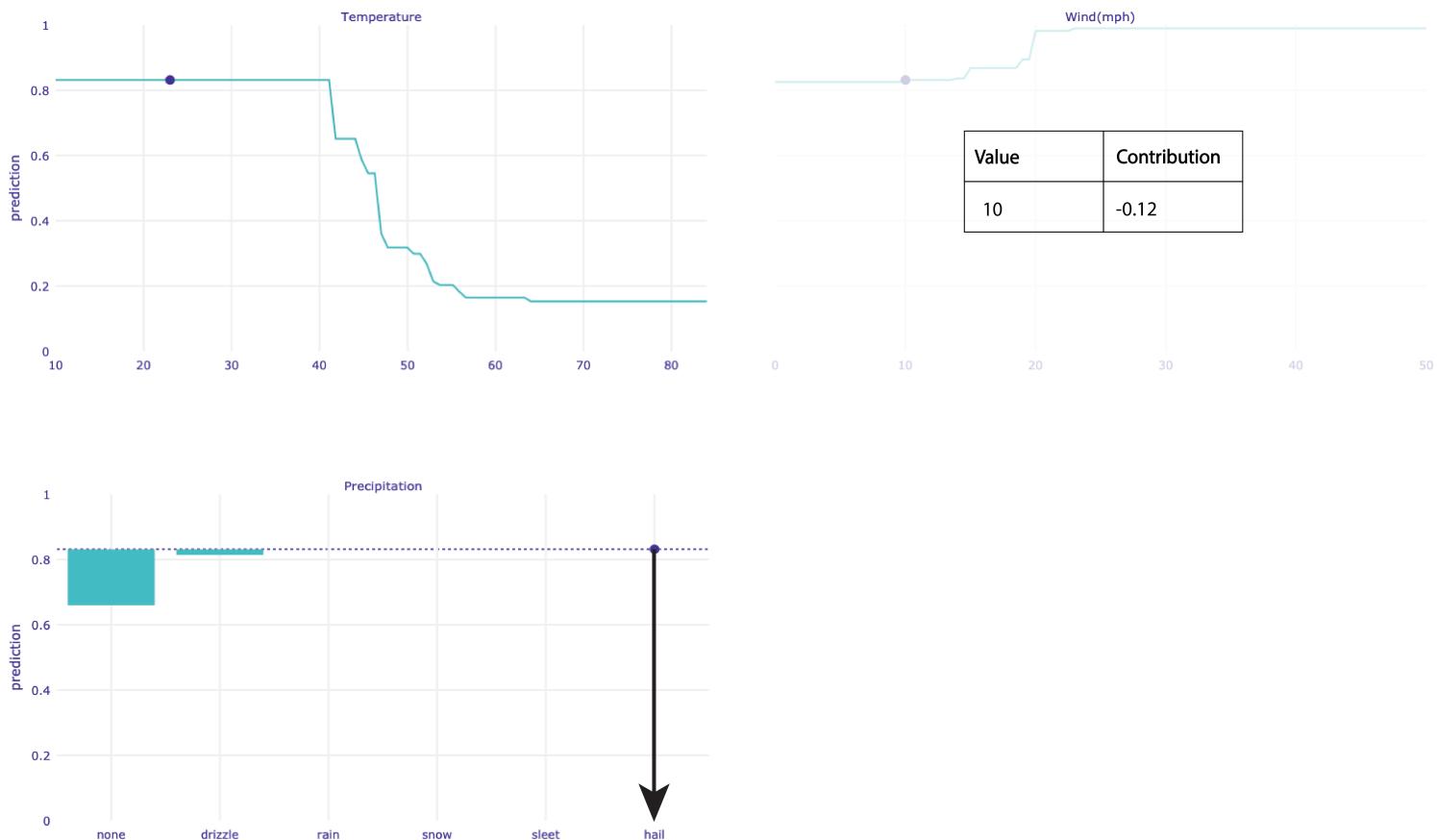
Model Prediction: 0.861



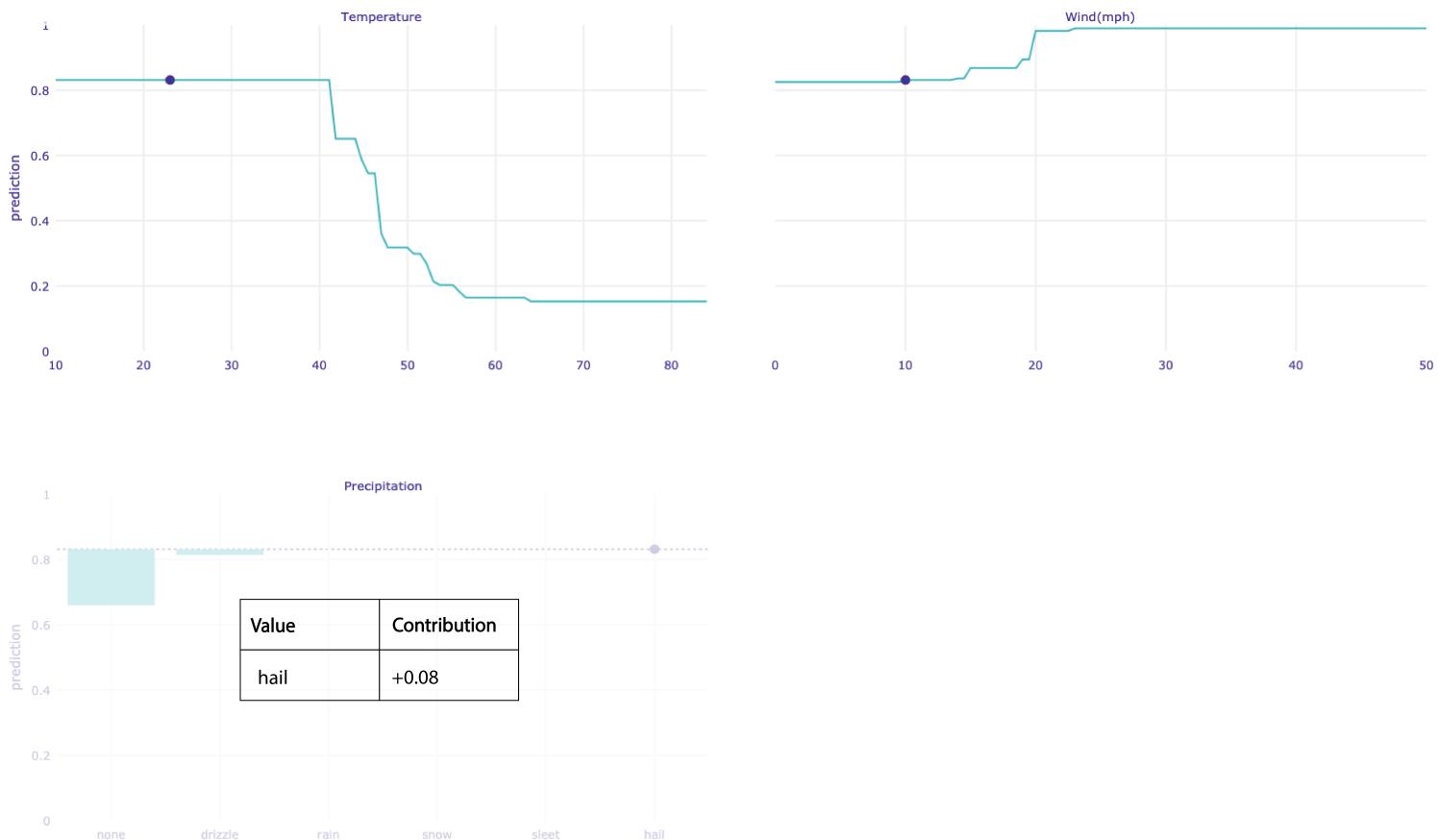
Model Prediction: 0.861



Model Prediction: 0.861



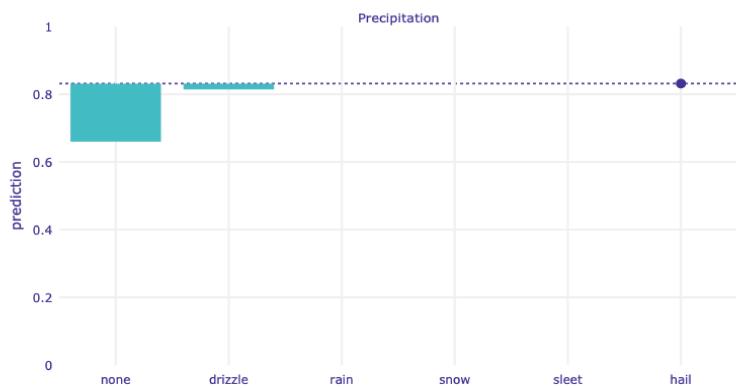
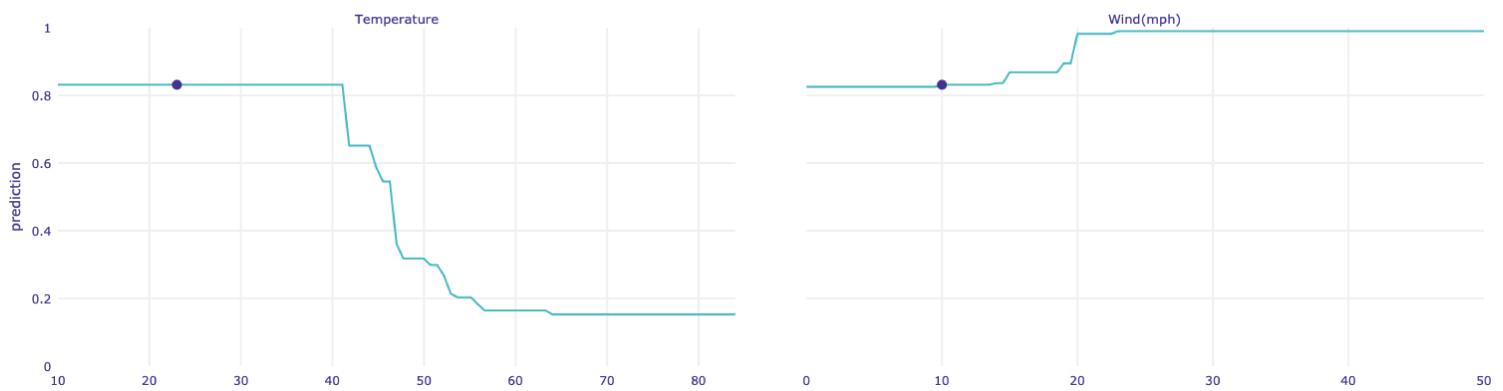
Model Prediction: 0.861



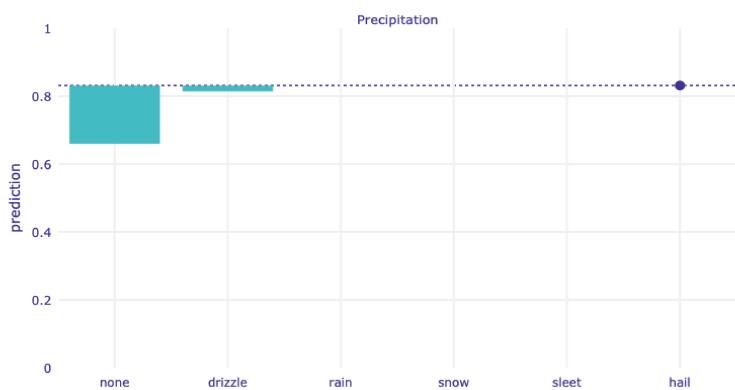
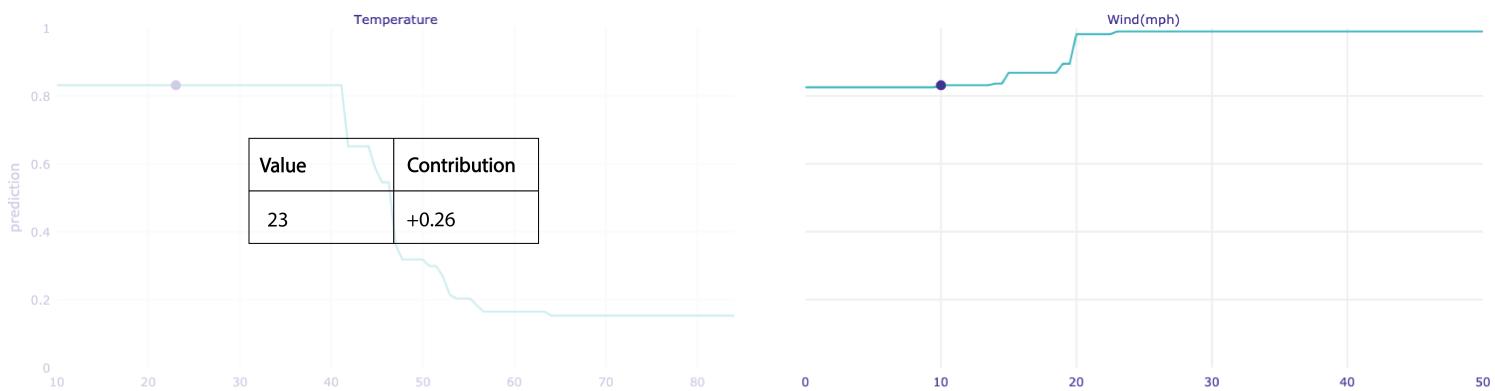
By looking at the explanation image, please select the value for **wind speed** input into the model:

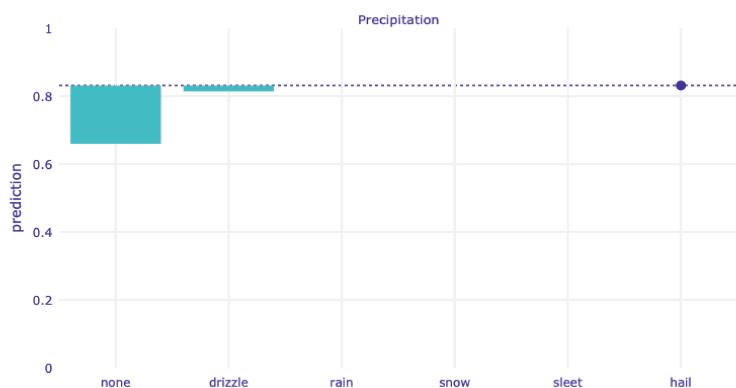
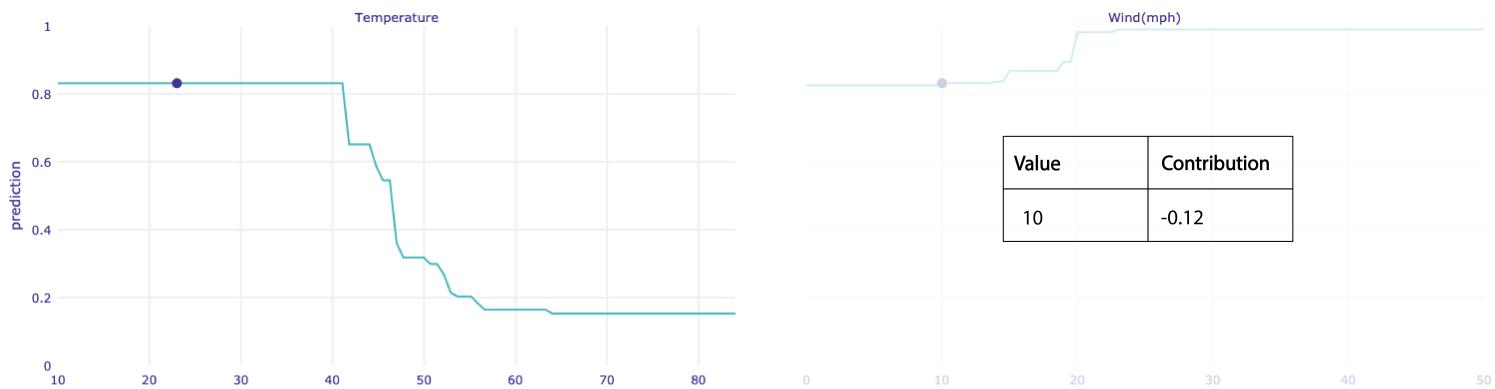
- 20 mph
- 0 mph
- 10 mph
- 5 mph
- 15 mph

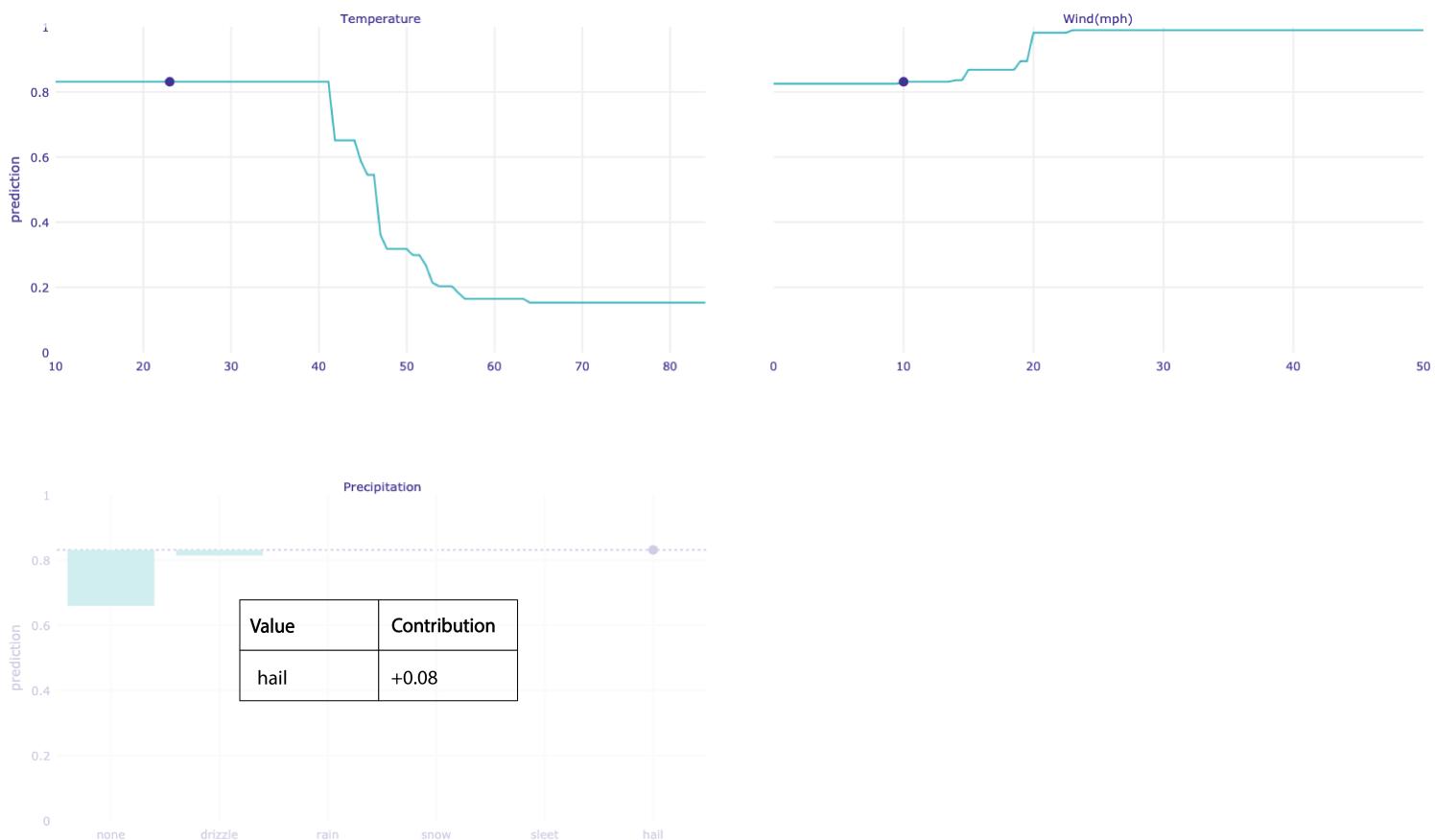
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861

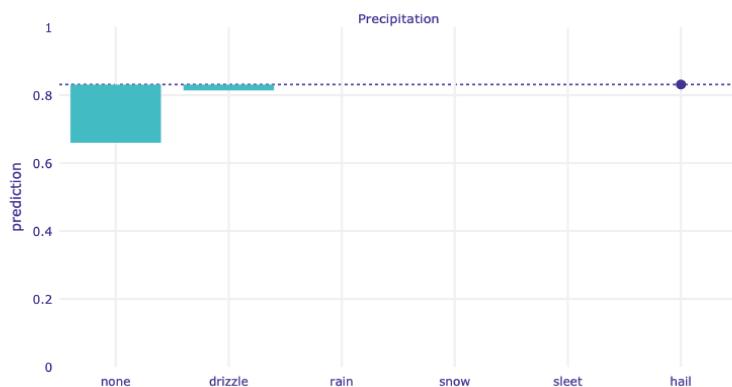
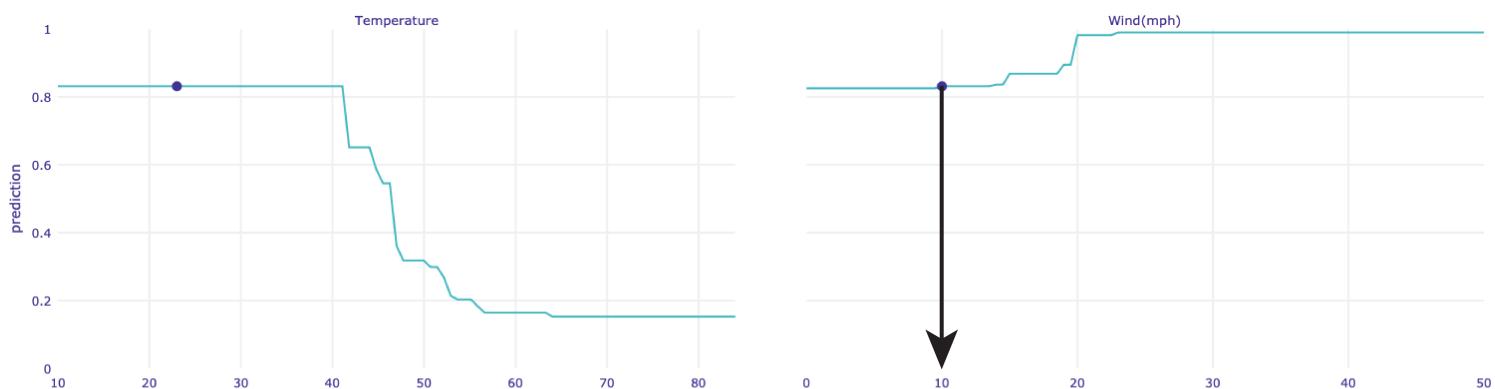
Model Prediction: 0.861

Correct – This value is in the **Value** column when you mouse over the Wind(mph) chart, and is below the blue dot in the Wind(mph) chart. This value is **10 mph**.

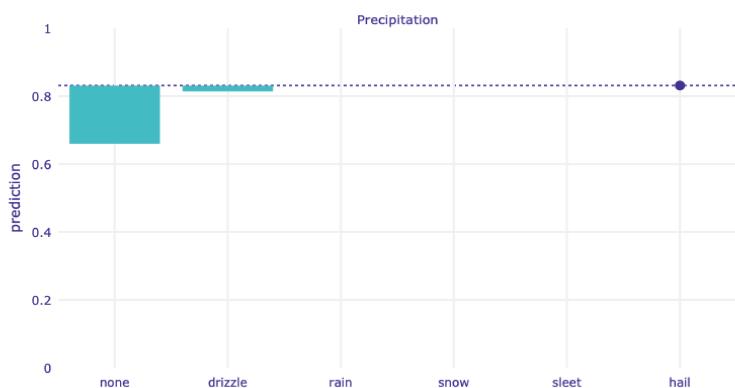
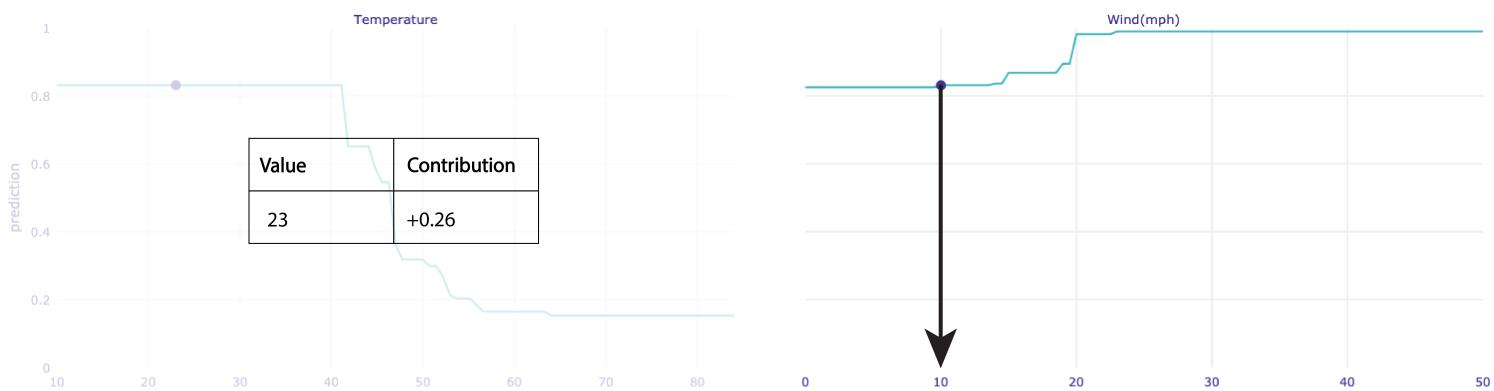
Not quite – This value is in the **Value** column when you mouse

over the Wind(mph) chart, and is below the blue dot in the Wind(mph) chart. This value is **10 mph**.

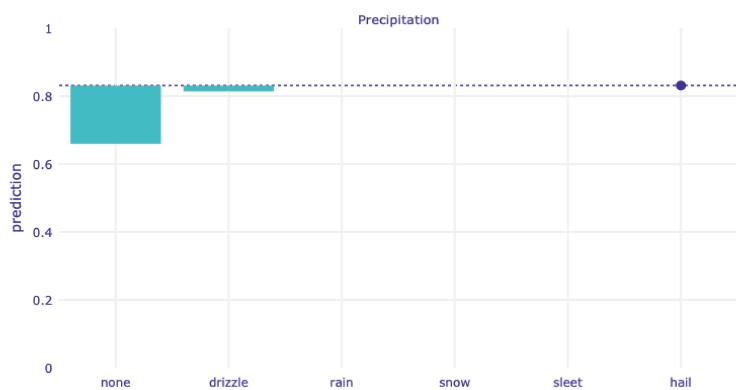
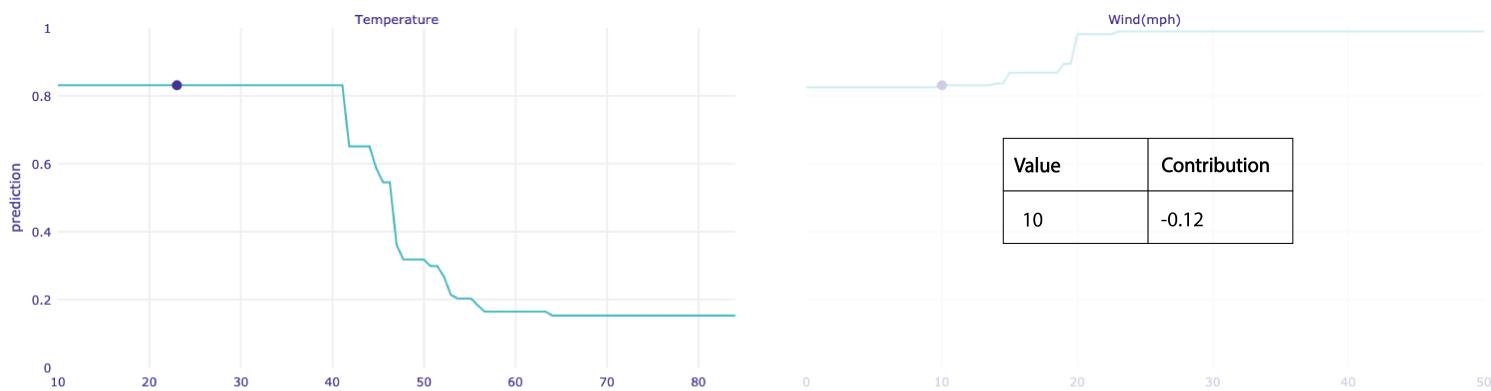
Model Prediction: 0.861



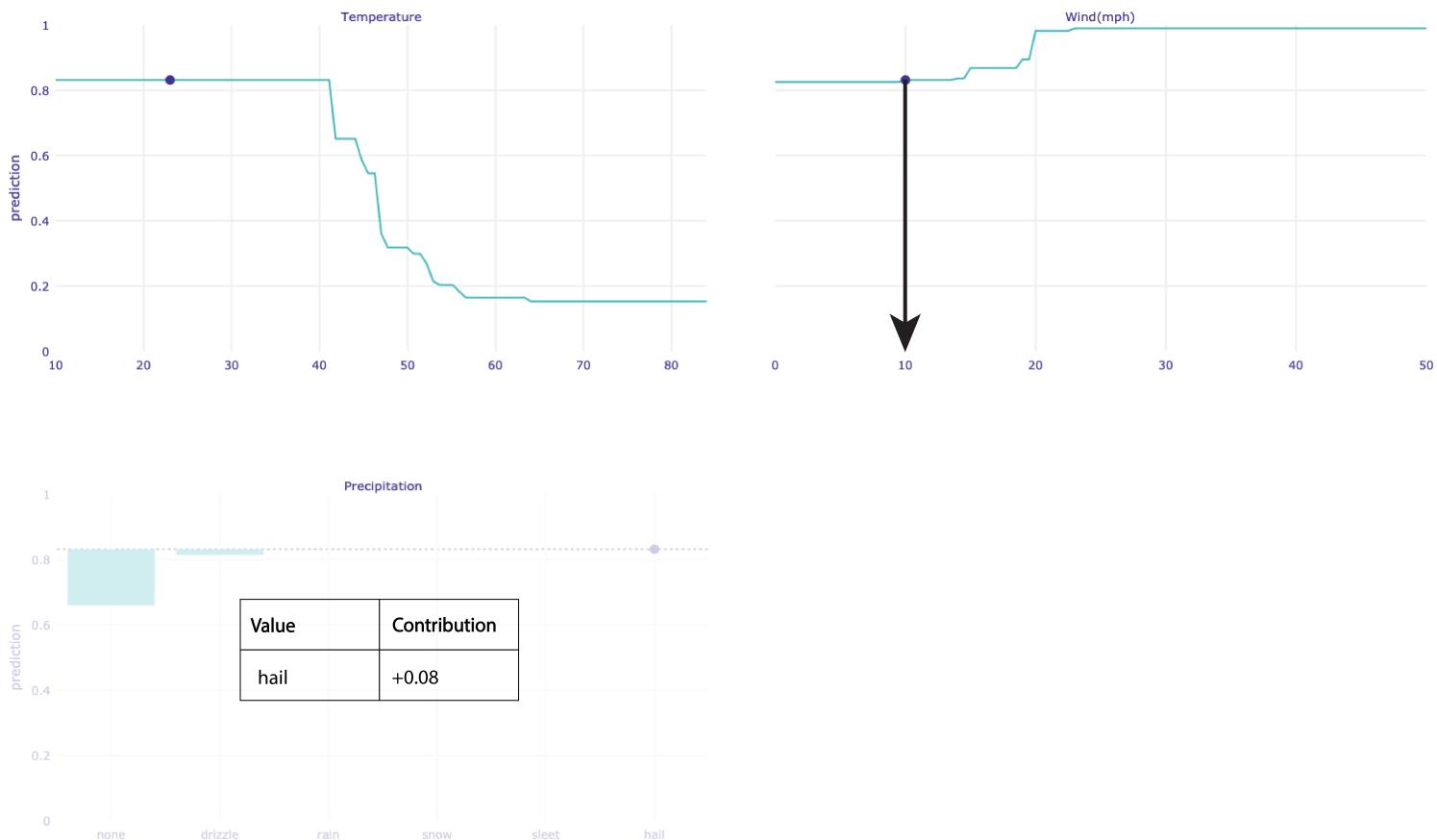
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861



Intro 8

All of the other values on these charts show you what the model's prediction would be, if the value of the corresponding feature was **changed** but the values of all of the other features were **NOT changed**.

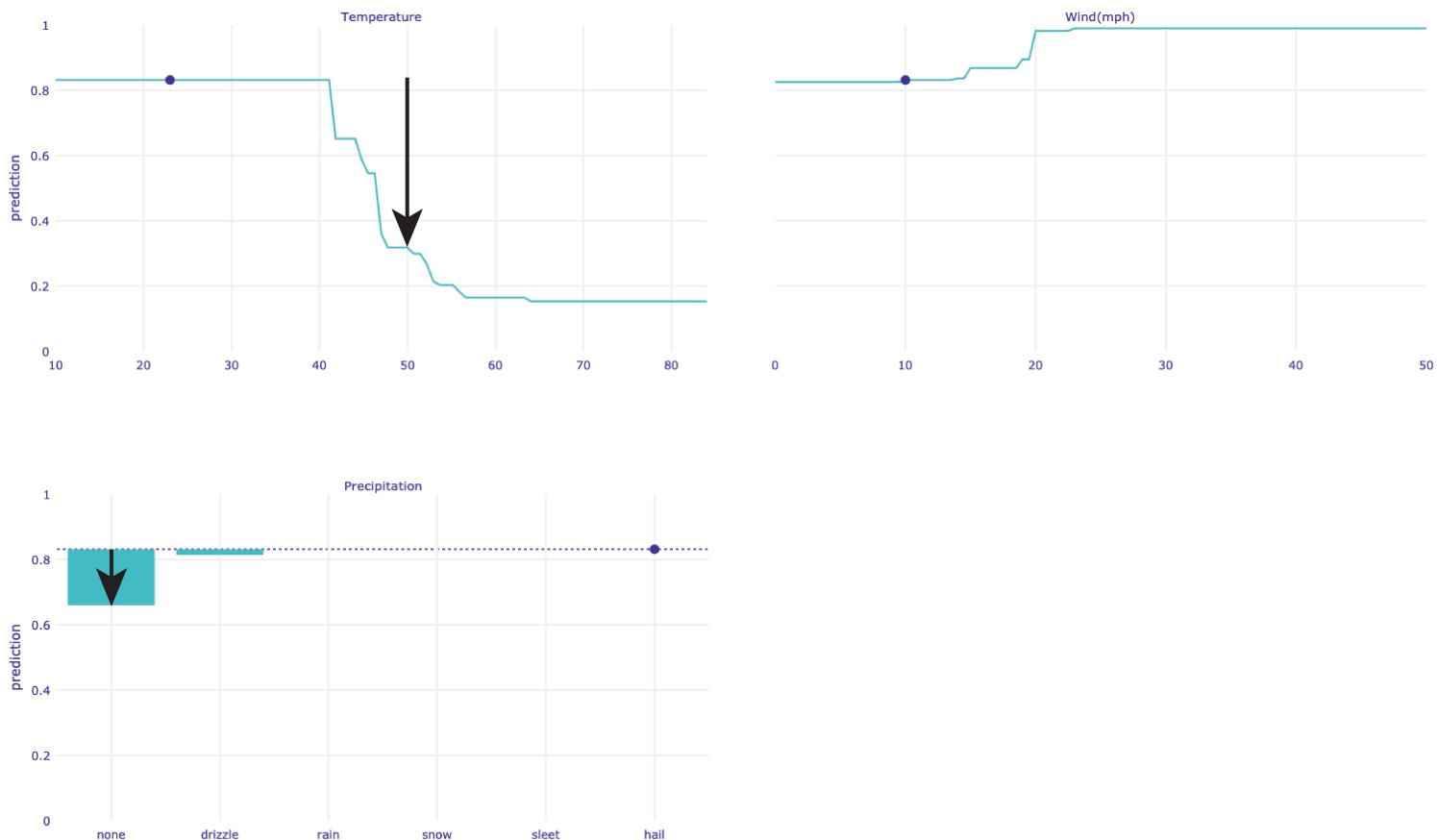
For example, in the Temperature chart below, you can see that if

the temperature was **50**, instead of **23**, then the model prediction would be about between **0.3** and **0.4**, meaning that it would now recommend you do NOT put on a coat.

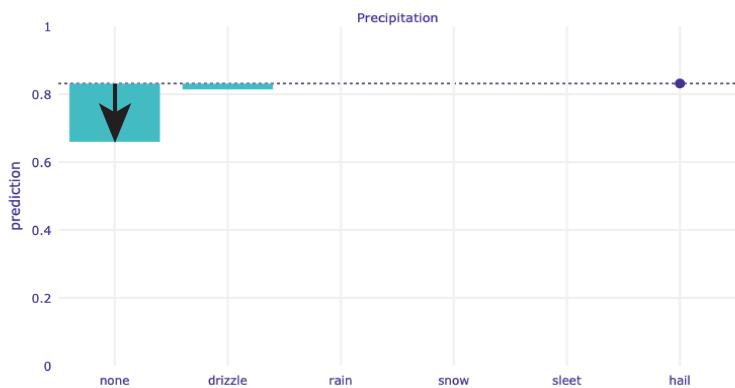
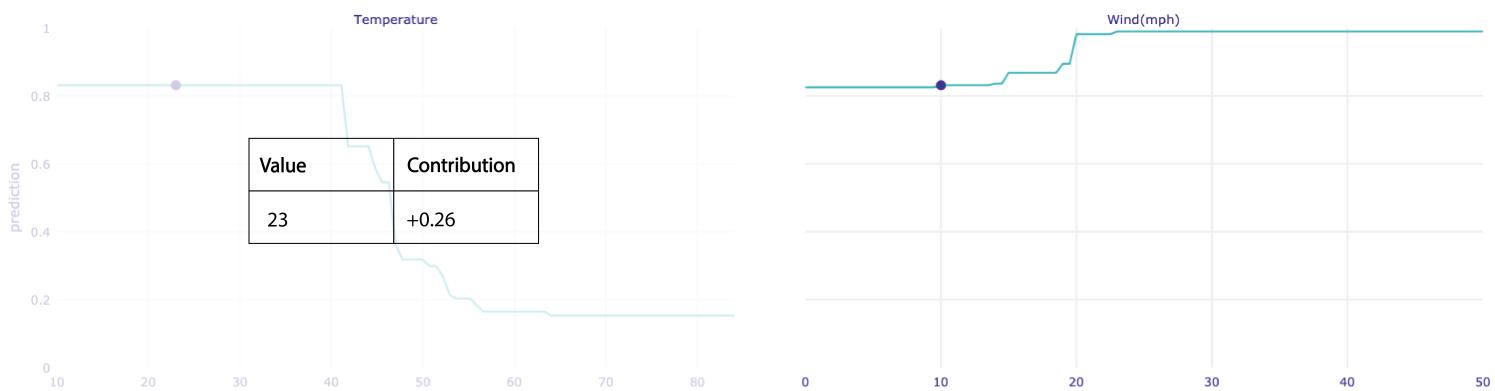
In the Precipitation bar chart below, you can see that if the precipitation was **None**, the model prediction would be between **0.6** and **0.7**. The model would still recommend that you put on a coat, but it would be less certain.

On the other hand, if the value of Temperature was instead 20, or the value of Precipitation was instead sleet, the model prediction would stay the same.

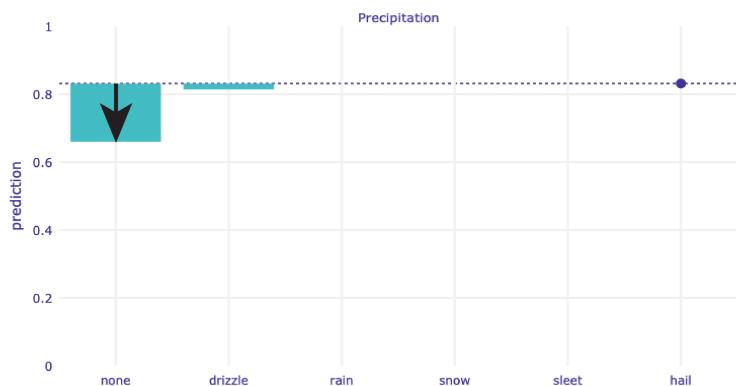
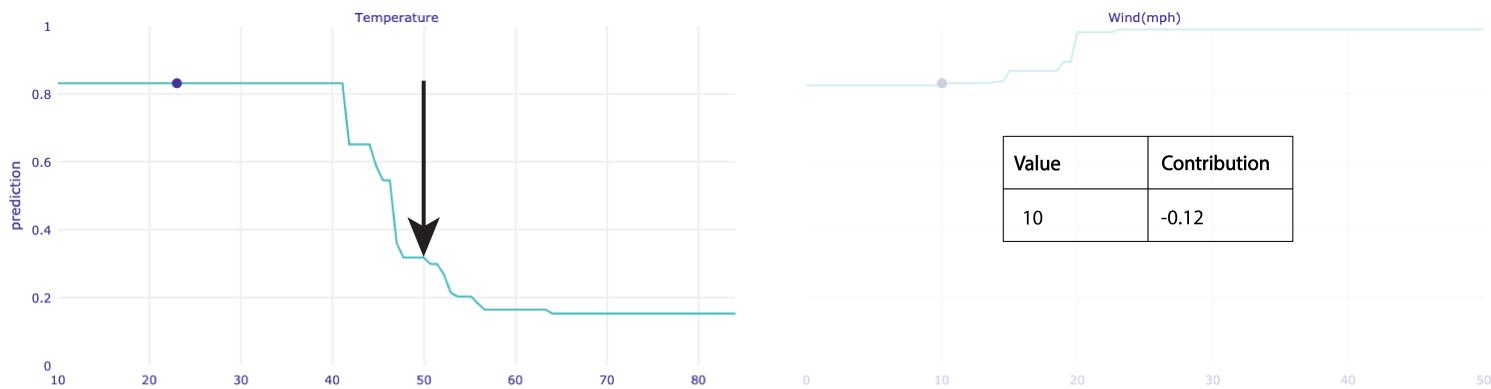
Model Prediction: 0.861



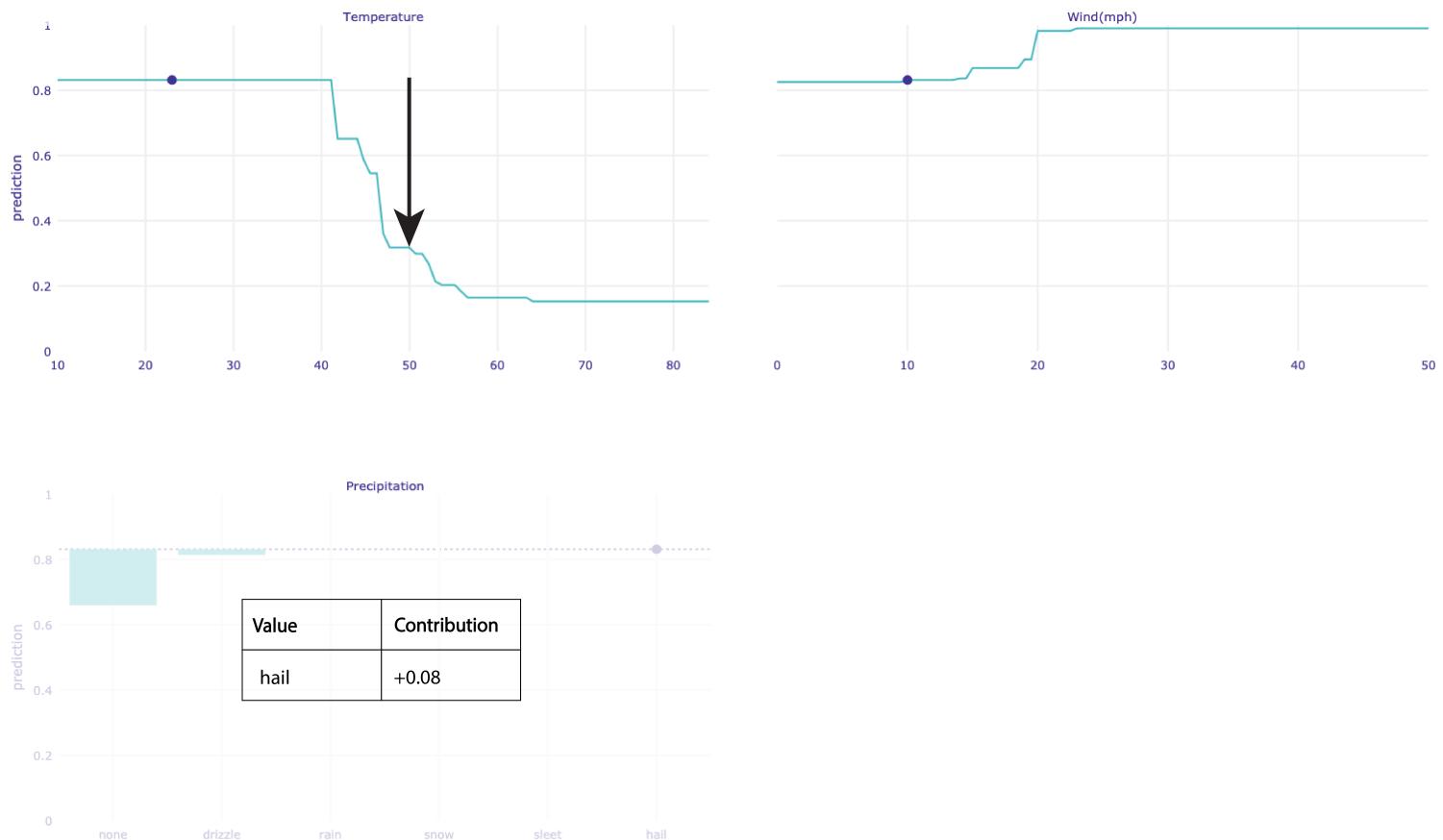
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861



Intro 9

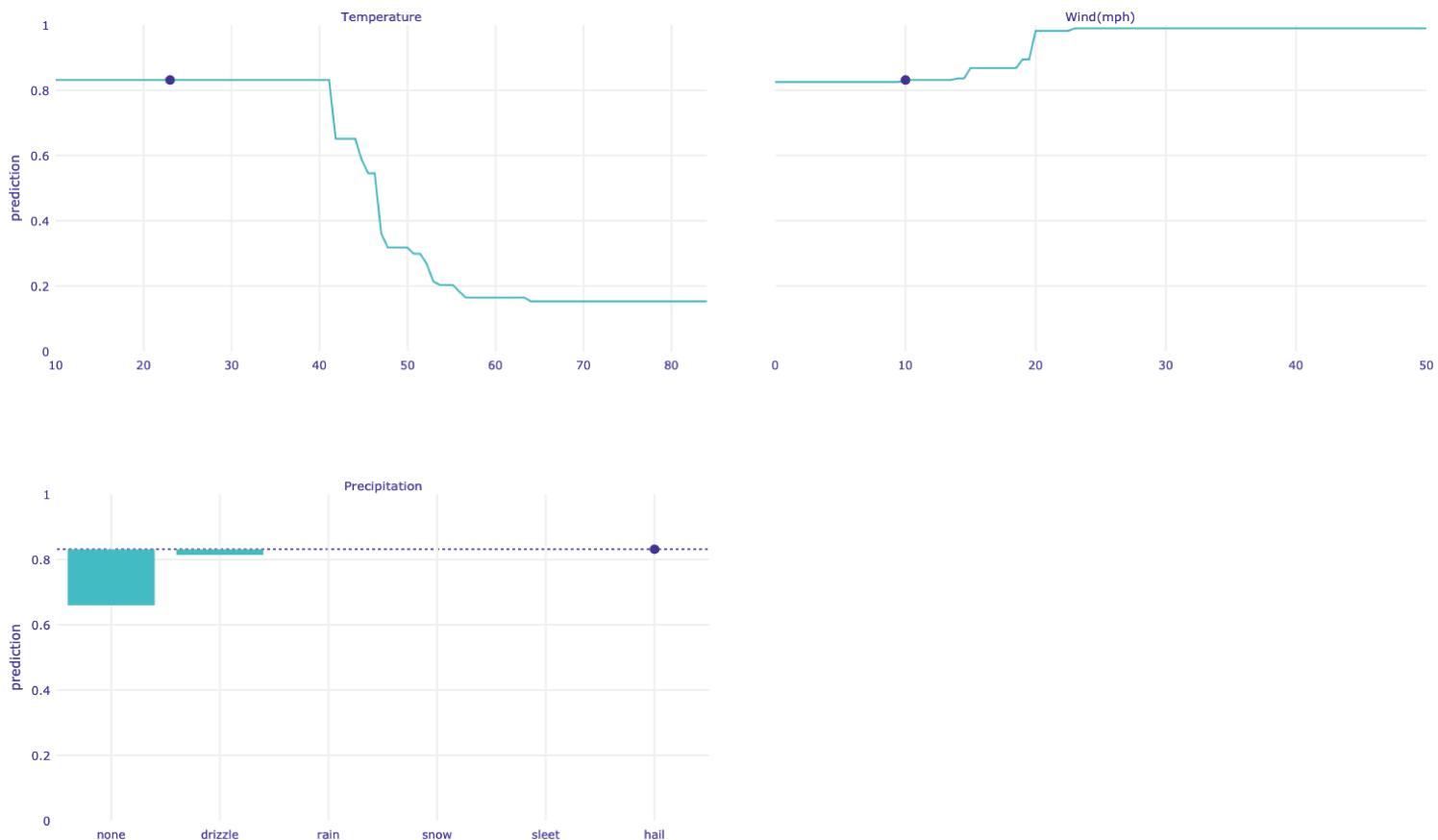
The predictive power of a feature, such as wind speed, depends on how much changing the feature's value will cause the output of the model to change. If a feature is important for this prediction, then changing its value will cause the prediction to change a lot.

To see the **exact** predictive power of a feature, mouse over its graph and look at the value in **Contribution** column of the table that you see.

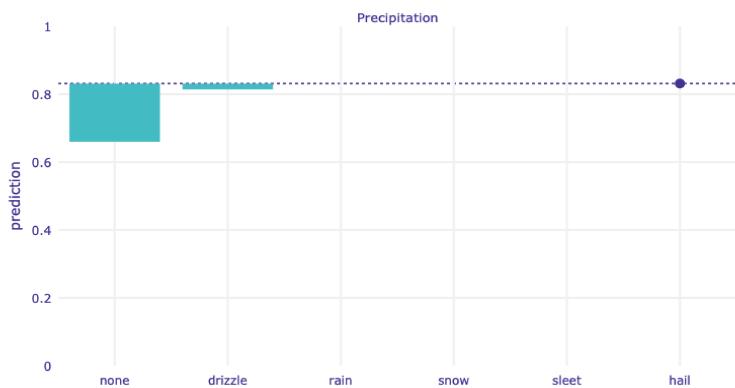
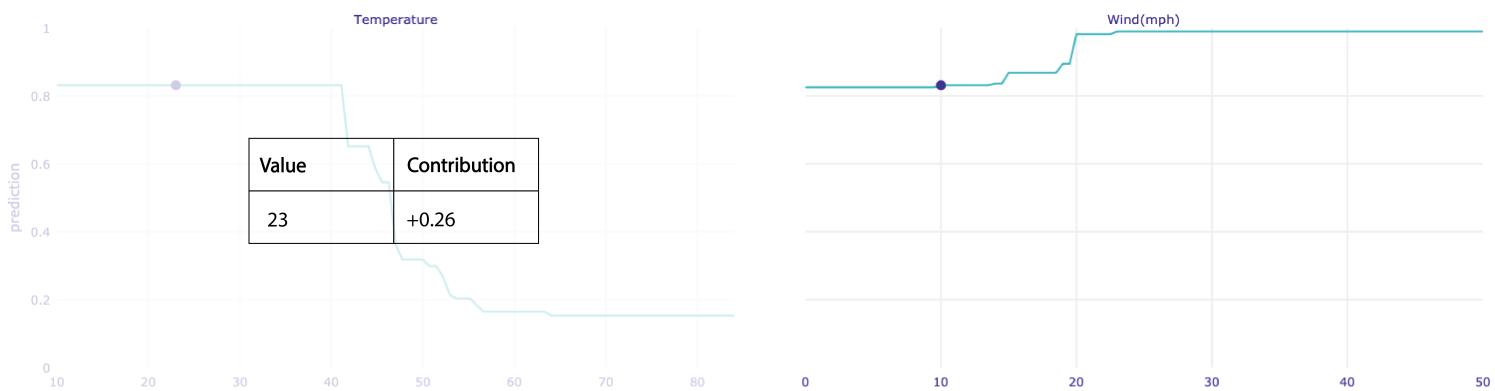
The feature that has the largest number in that column (+ and - signs **don't** matter here) has the largest predictive power.

Here, you can see Temperature has the largest predictive power, because 0.26 is larger than 0.12 (Wind) or 0.08 (Precipitation).

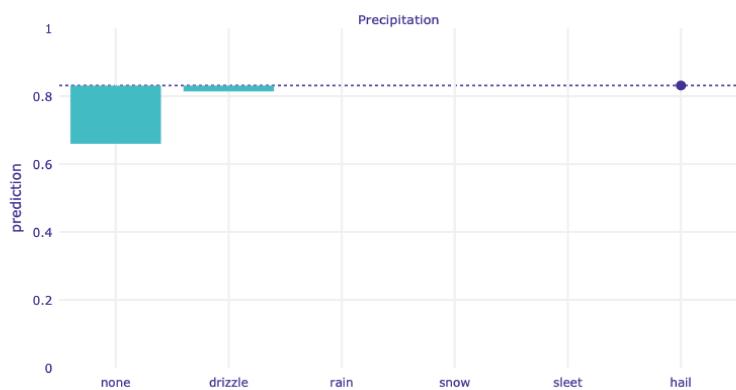
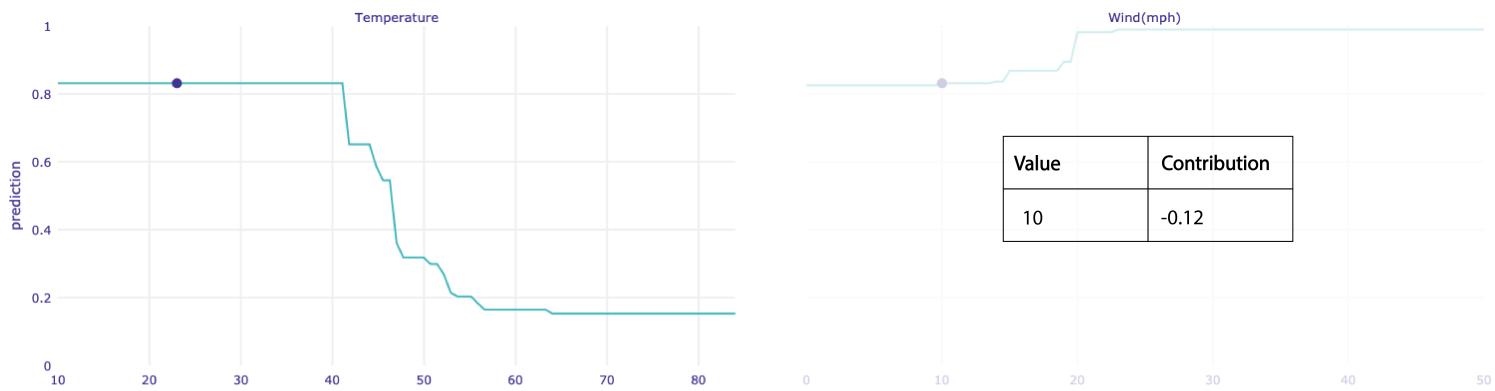
Model Prediction: 0.861



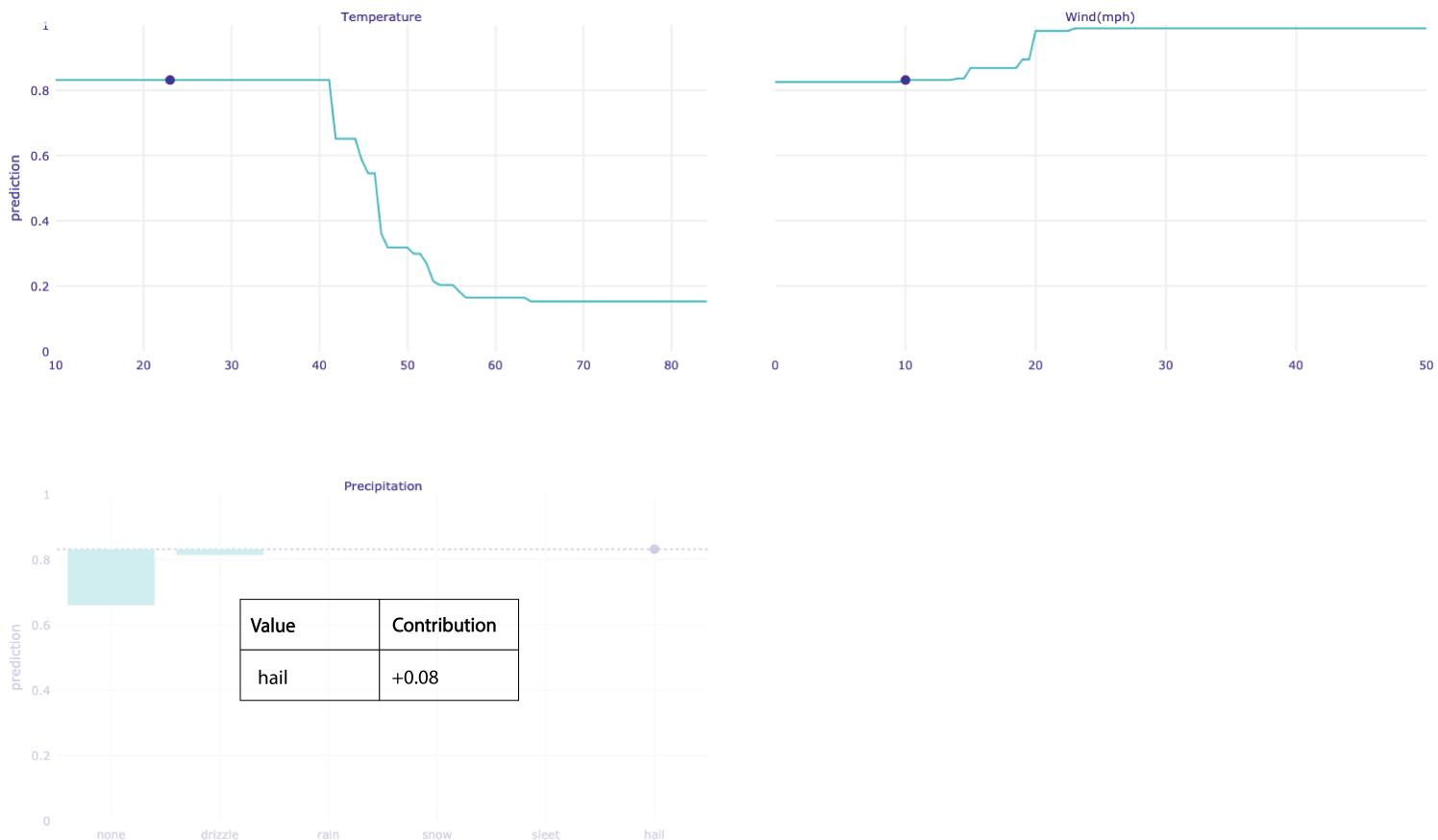
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861



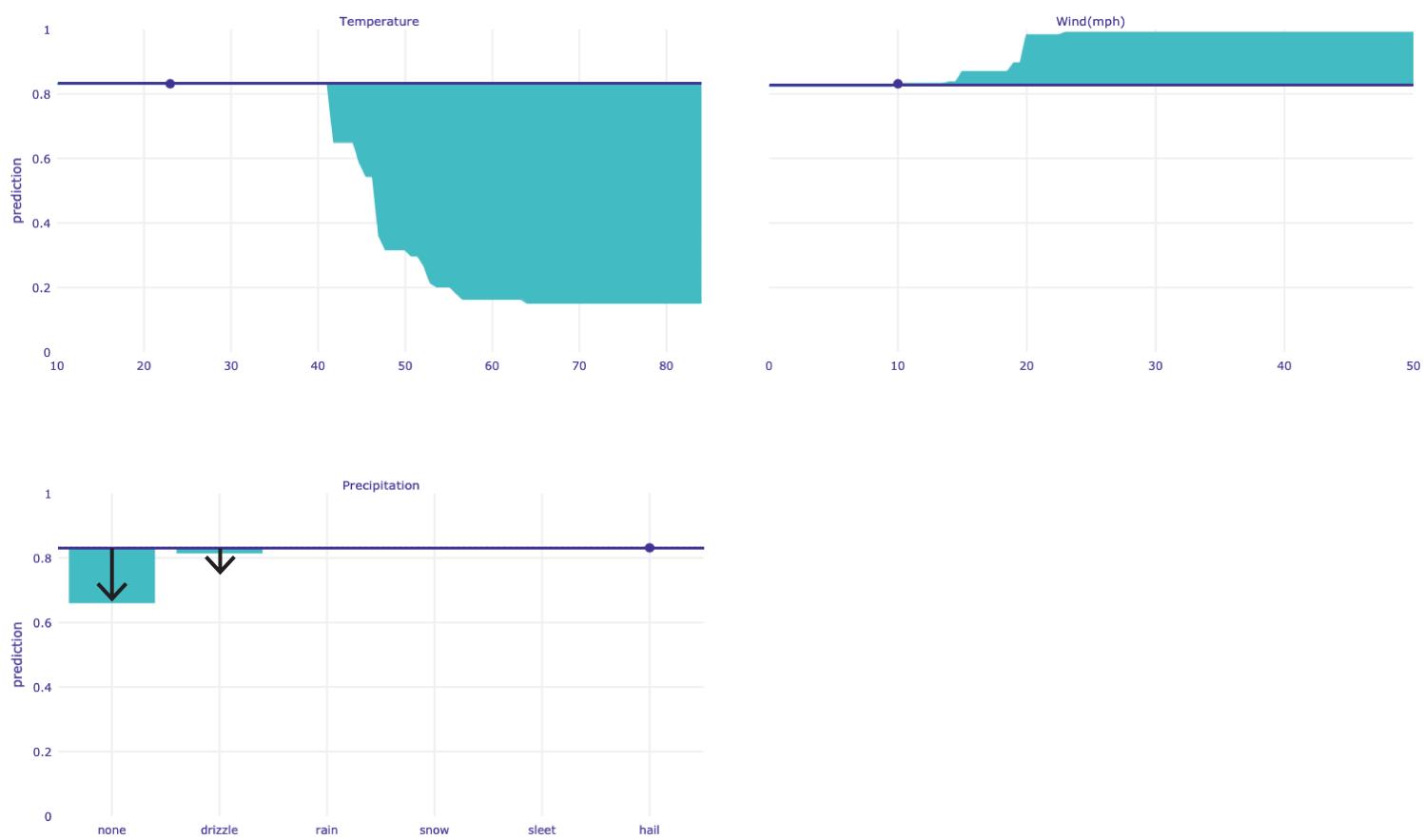
Intro 10

You can also get an idea of the predictive power of a numerical feature by eyeballing the area between the line that you see, and an imaginary horizontal line at the current prediction value, where the blue dot is.

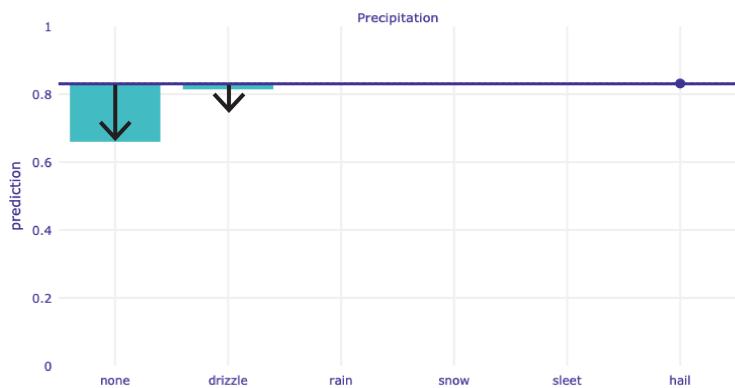
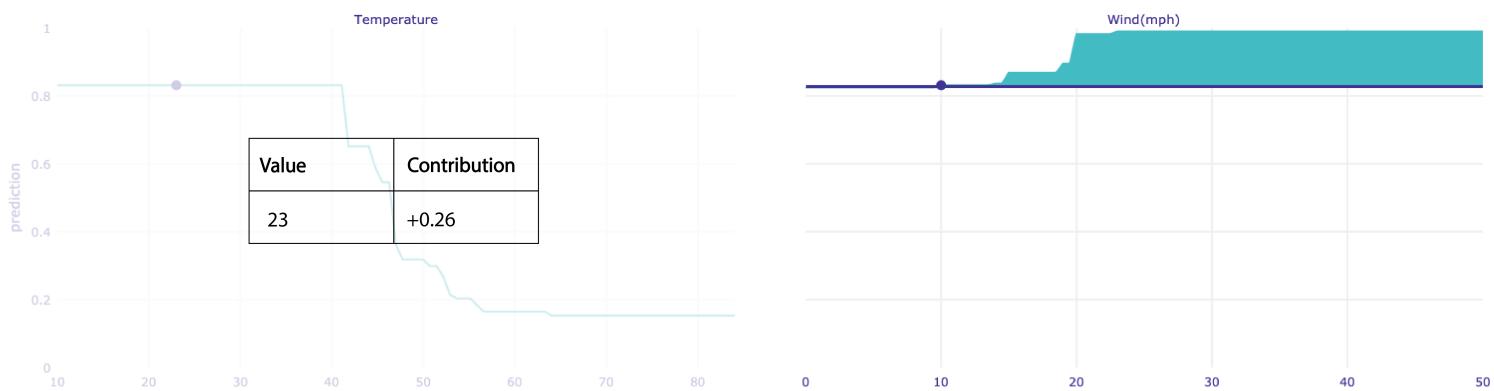
For bar charts, eyeball the sizes of the bars.

Here, you can see that temperature has more predictive power than wind, because a change in temperature changes the model output more than a change in wind does, and the area between the line and the current prediction value is larger.

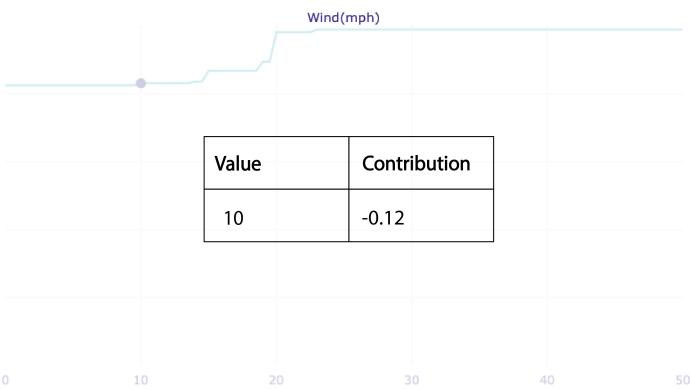
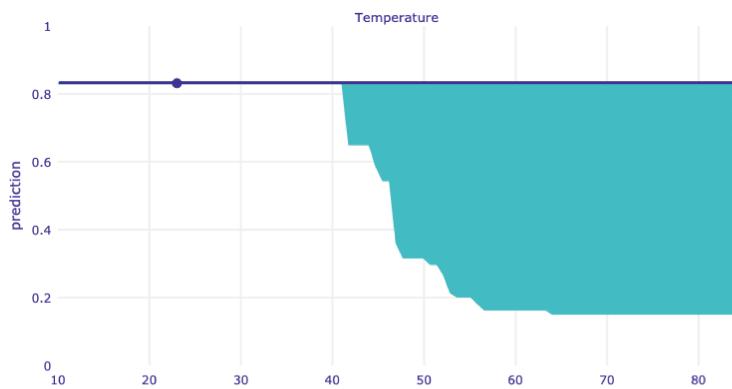
Model Prediction: 0.861



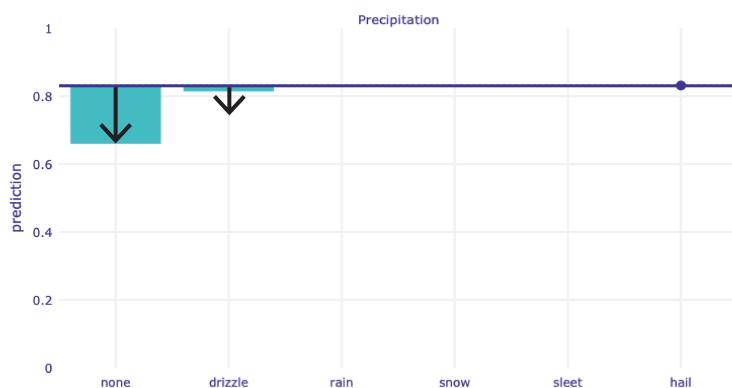
Model Prediction: 0.861



Model Prediction: 0.861



Value	Contribution
10	-0.12



Model Prediction: 0.861

Intro 11

Each input value of temperature, wind, and precipitation can push the model's prediction to be higher or lower. If the final prediction is pushed above 0.5, the model will return 'YES' (wear a coat). If the final prediction is pushed below 0.5, the model will return 'NO' (do not wear a coat).

To tell whether a feature value is pushing the model toward predicting "YES" or "NO", look at the **sign** of the value in the **Contribution** column of the table that you see when you mouse over the feature graph.

You can ignore the numbers for this.

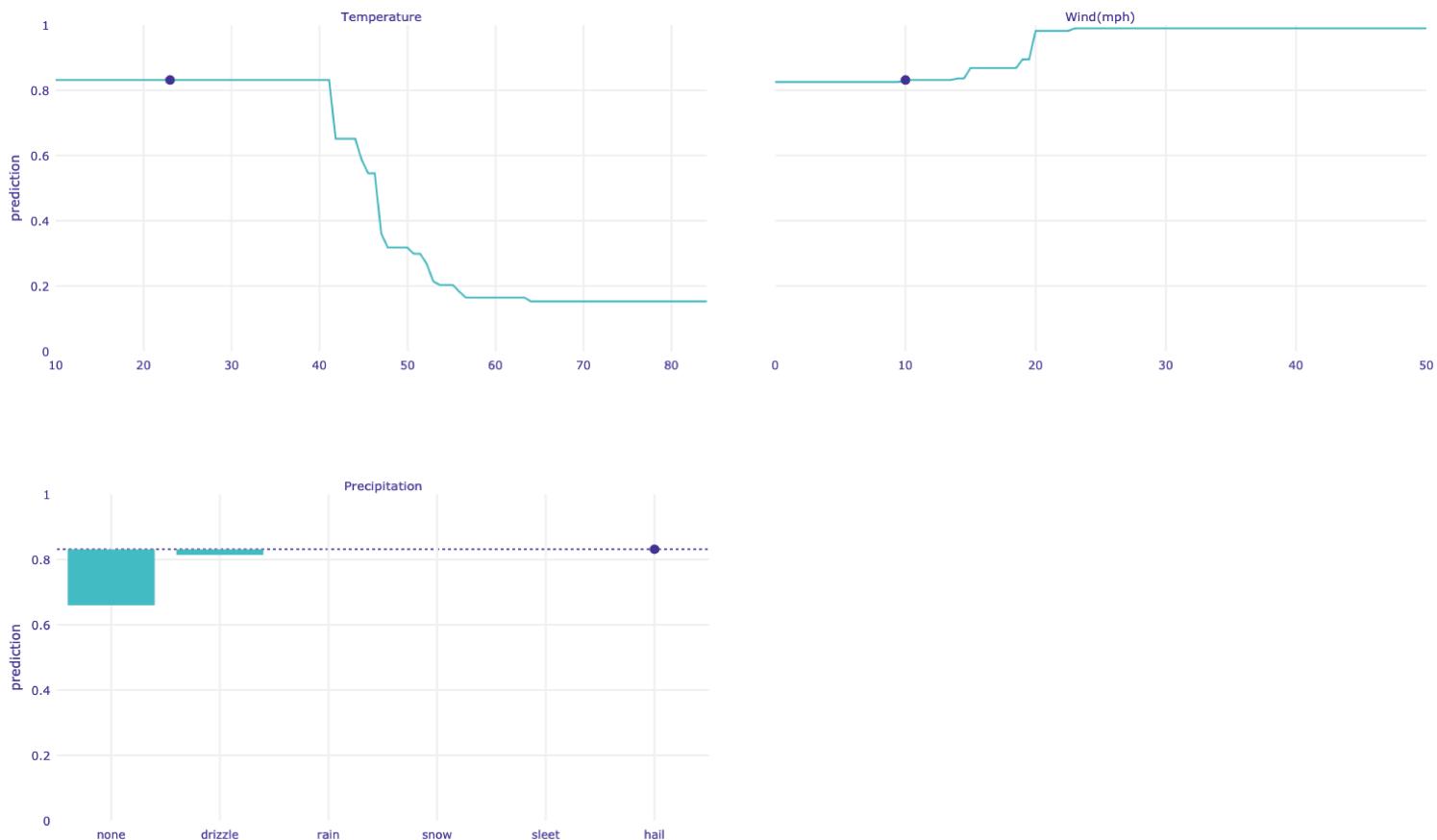
Features contributions with **positive signs** (+) push the model toward predicting "YES".

Features contributions with **negative signs** (-) push the model toward predicting "NO".

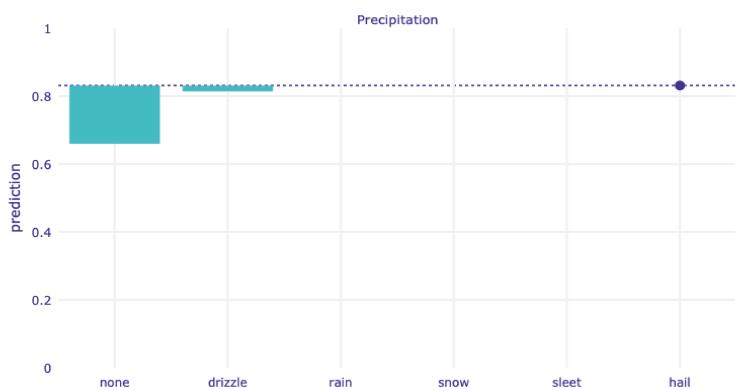
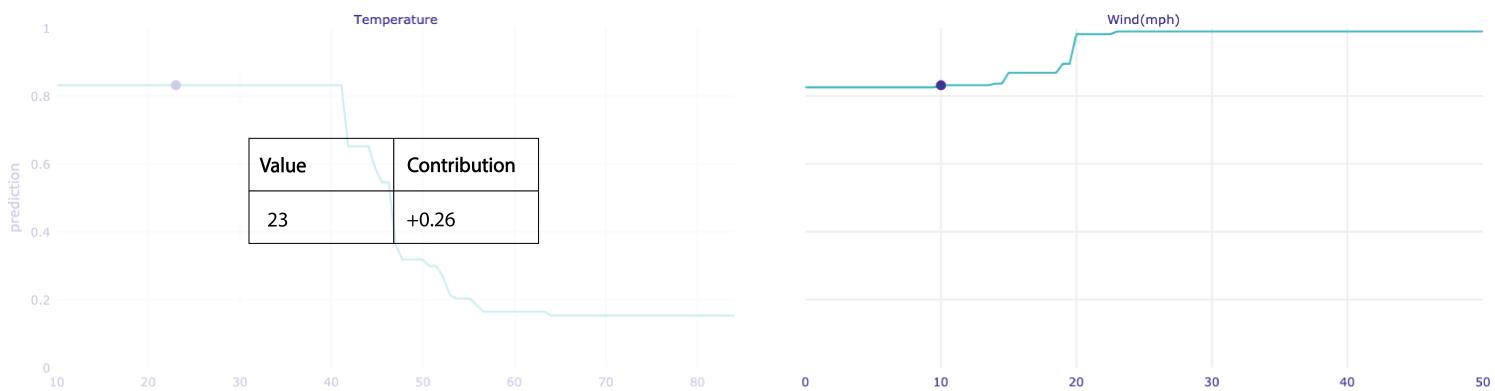
Here, temperature and precipitation contributions have positive (+) signs, so the values of these features are pushing the model toward "YES".

Wind's contribution has a negative (-) sign, so the value of this feature is pushing the model toward predicting "NO".

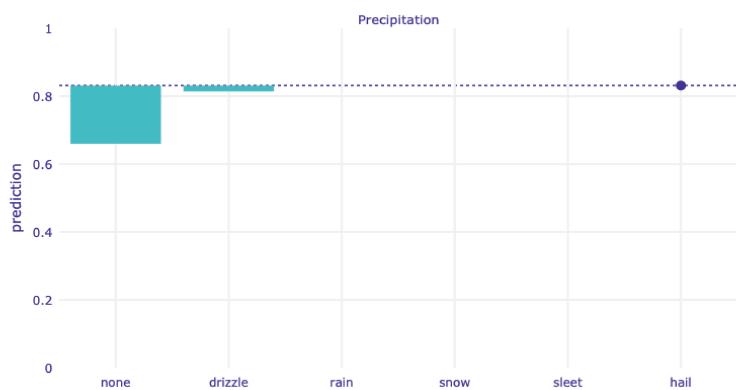
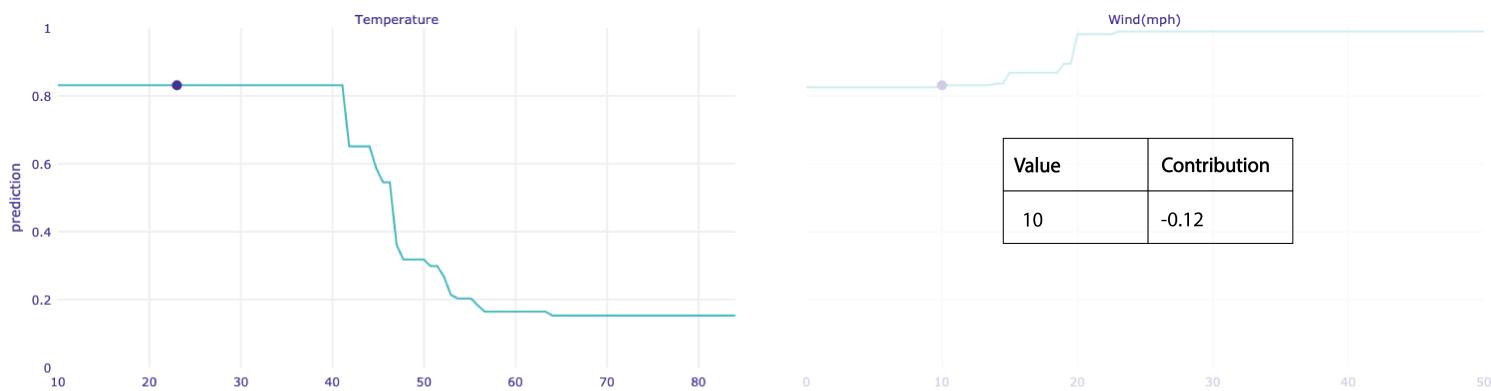
Model Prediction: 0.861



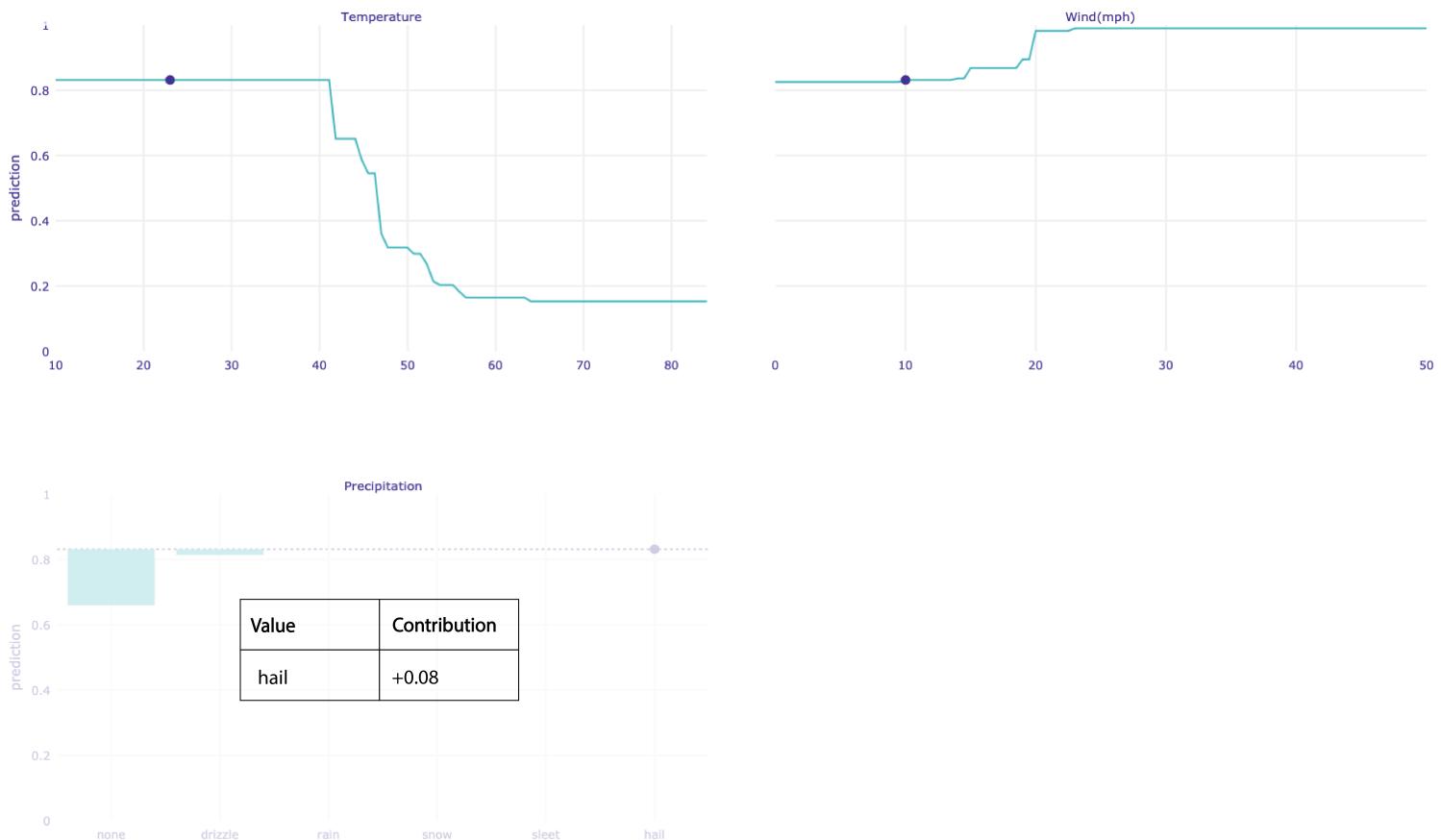
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861



Intro 12

Another way to tell that a feature value is pushing the model toward predicting 'YES' if changing this value is more likely to **decrease** the model prediction than increase it. For a numerical feature, this means that the area **below** an imaginary horizontal line at the current prediction value dot is **larger** than the area above the line. For a categorical feature, this means that the sum

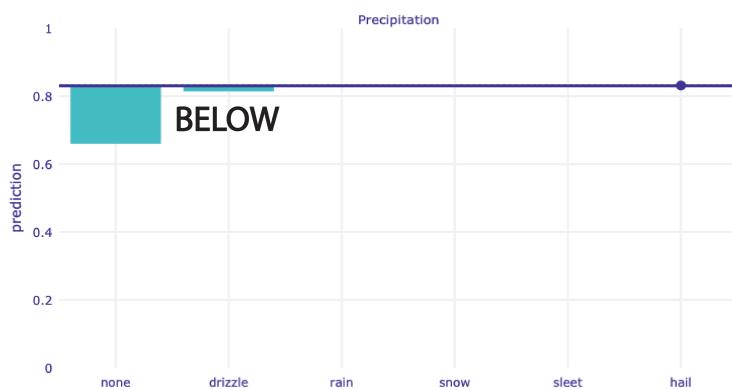
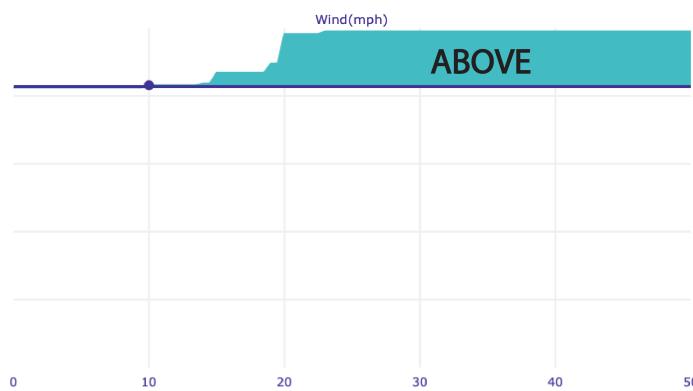
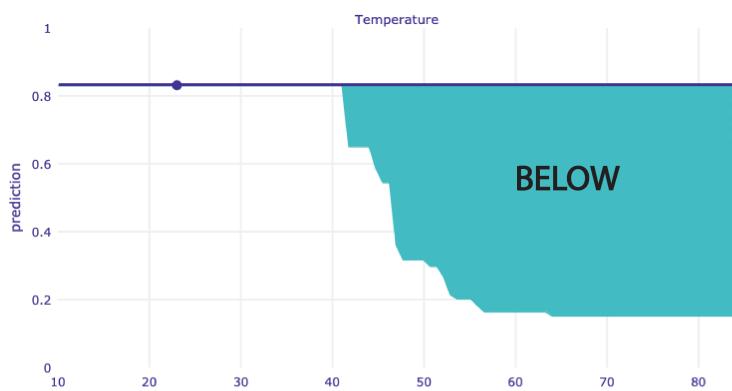
of the bars **below** the blue dot is larger than the sum of the bars **above** the blue dot.

For instance, in this example, the area below the current prediction dot in the Temperature chart is large, while the area above the prediction dot is 0. Changing the value of the temperature will probably **decrease** the probability that you should wear a coat. This means the current value of temperature is pushing the model toward predicting 'YES'.

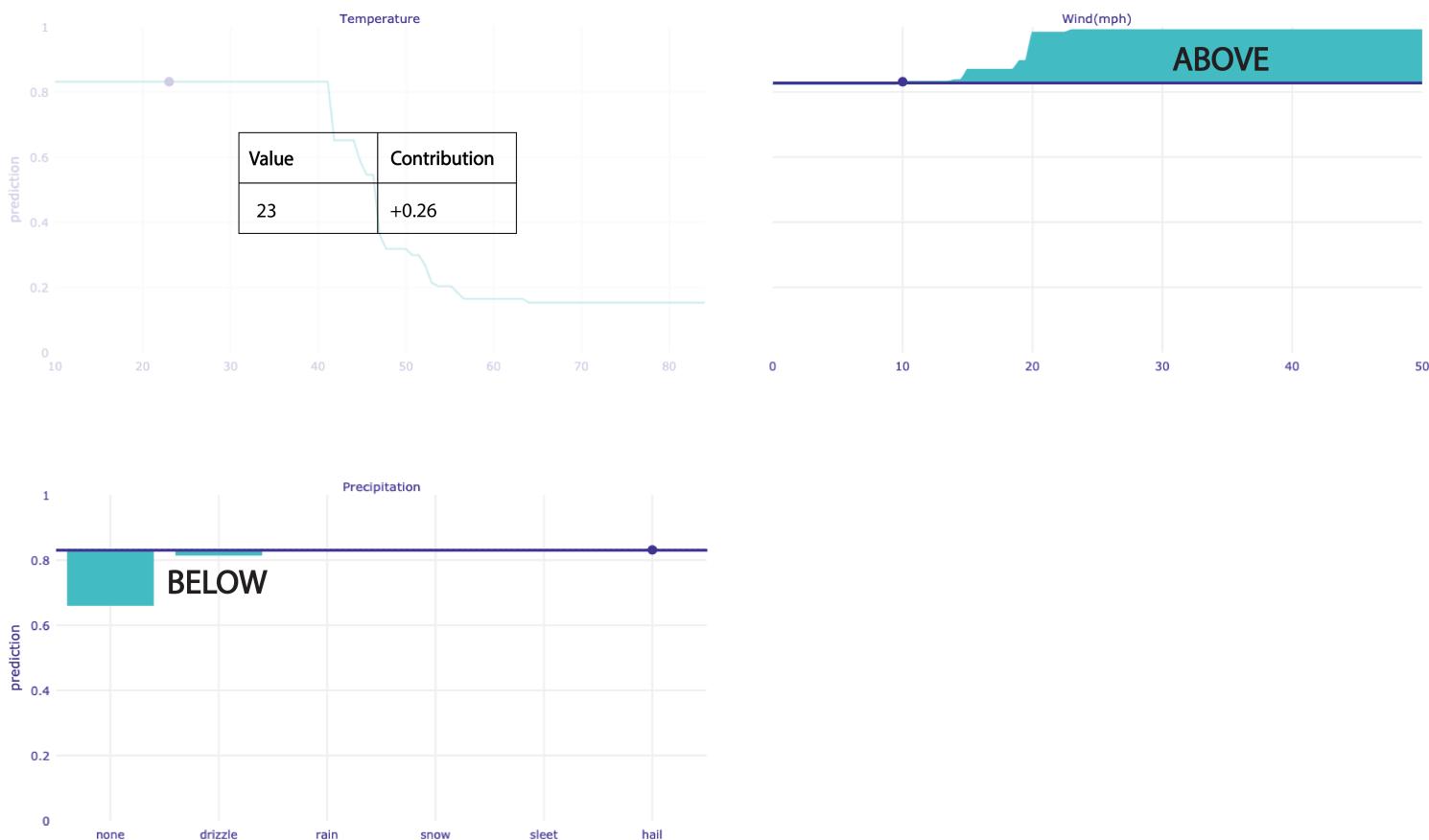
You can tell that a feature value is pushing the model toward predicting 'NO' if changing this value is more likely to **increase** the model prediction than decrease it. For a numerical feature, that means that the area **above** an imaginary horizontal line at the current prediction value is **larger** than the area below the line. For a categorical feature, this means that the sum of the bars **above** the blue dot is larger than the sum of the bars **below** the blue dot.

The area above the current prediction dot in the Wind(mph) chart is medium size, while the area below the prediction dot is 0. Changing the value of wind(mph) will probably **increase** the probability that you should wear a coat. This means the current value of wind(mph) is pushing the model toward predicting 'NO'.

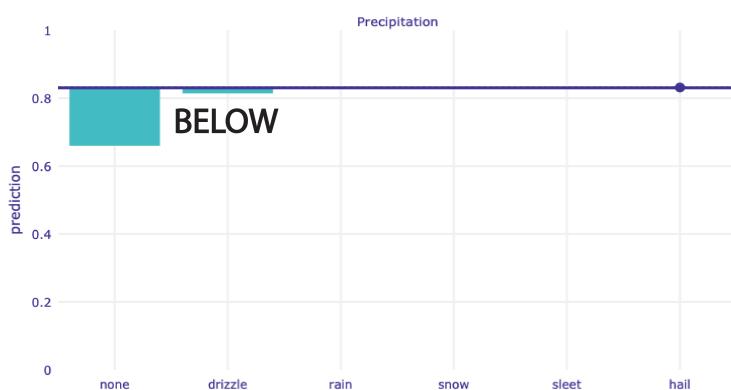
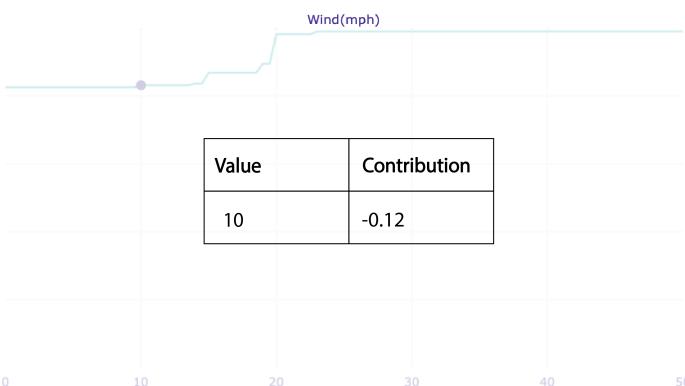
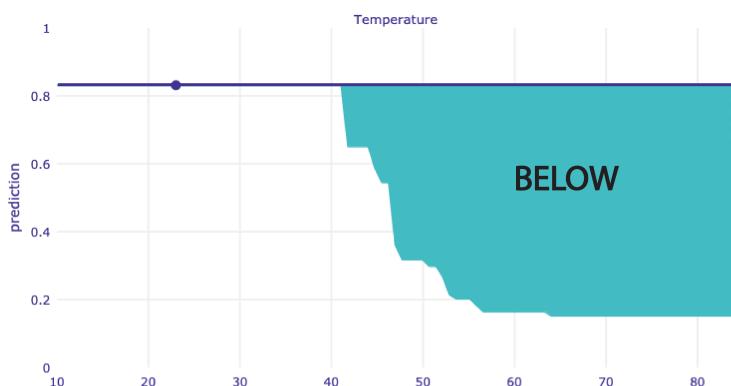
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861

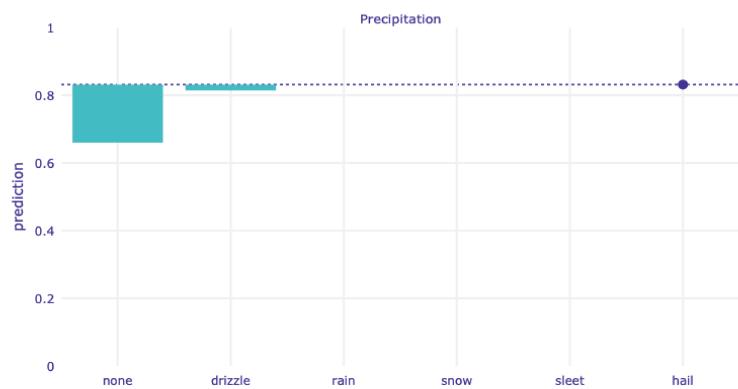
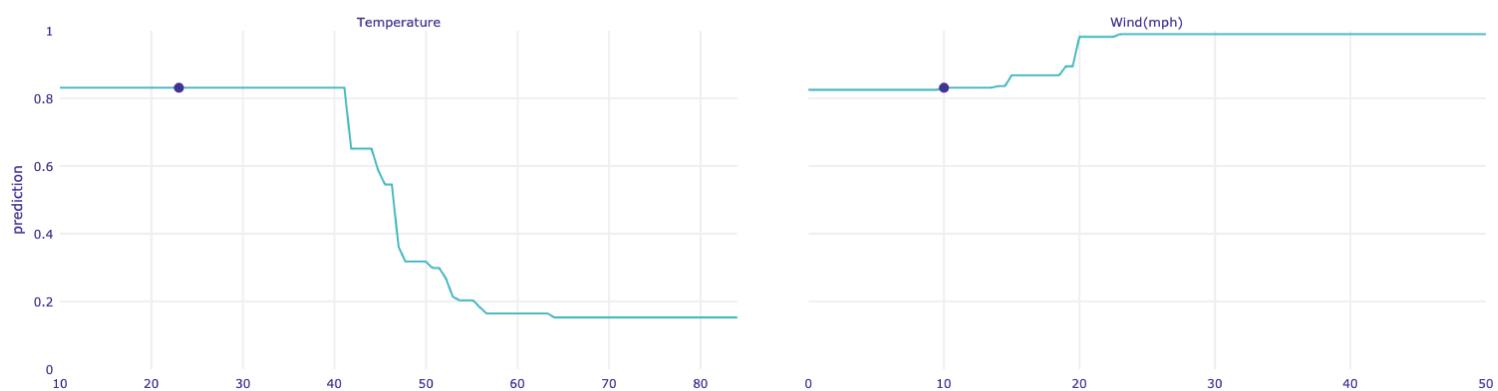


Intro Test 3

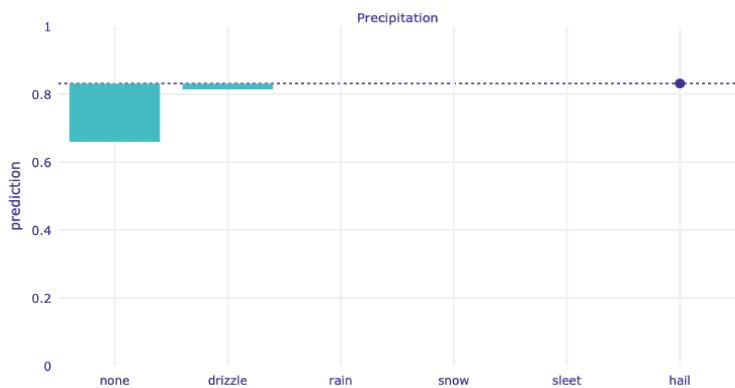
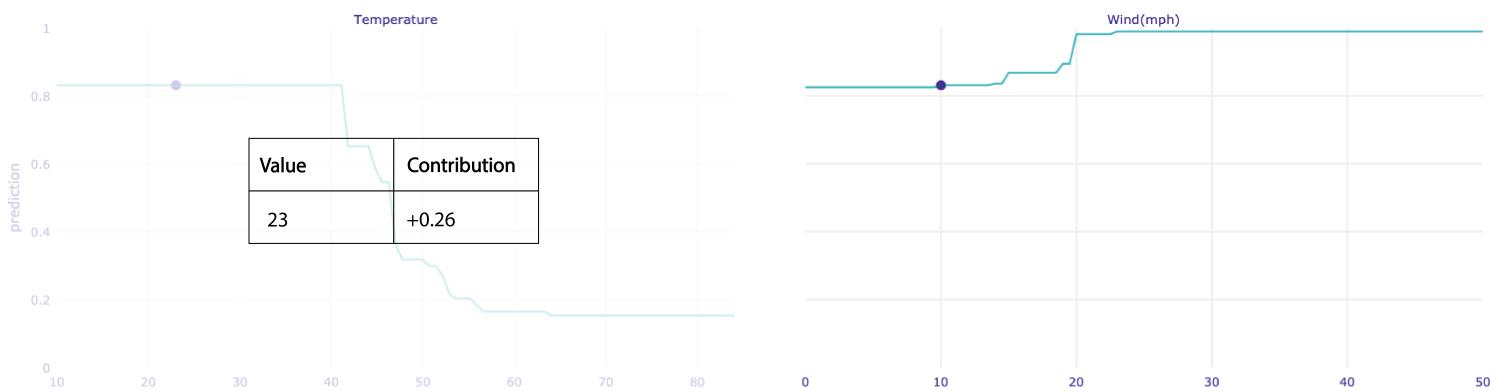
As a review, by looking at the explanation image, which factor(s) are pushing the model toward predicting 'YES'?

- Temperature
- Wind
- Precipitation

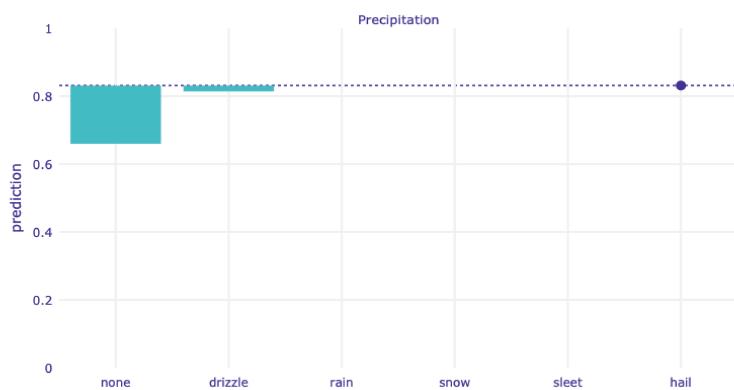
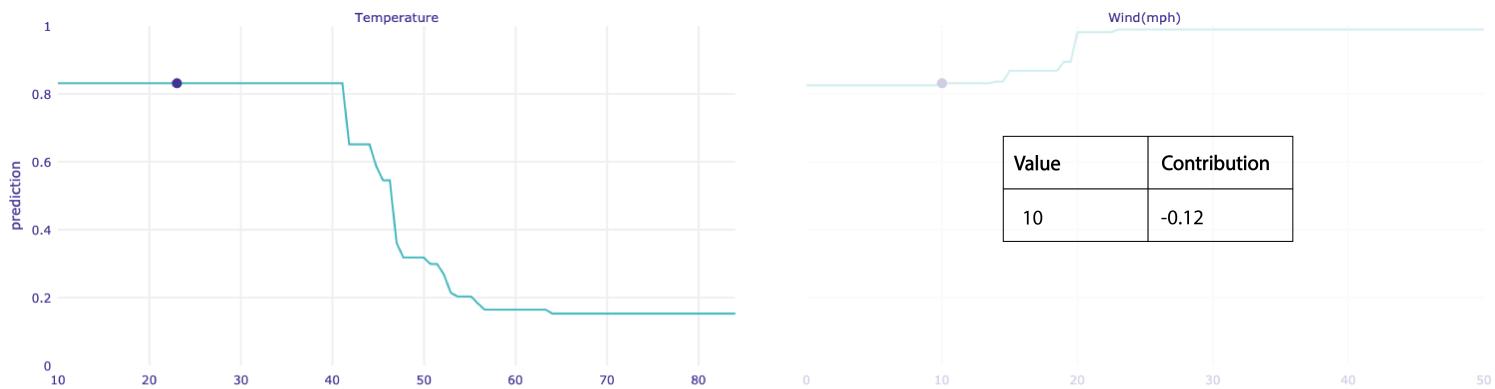
Model Prediction: 0.861



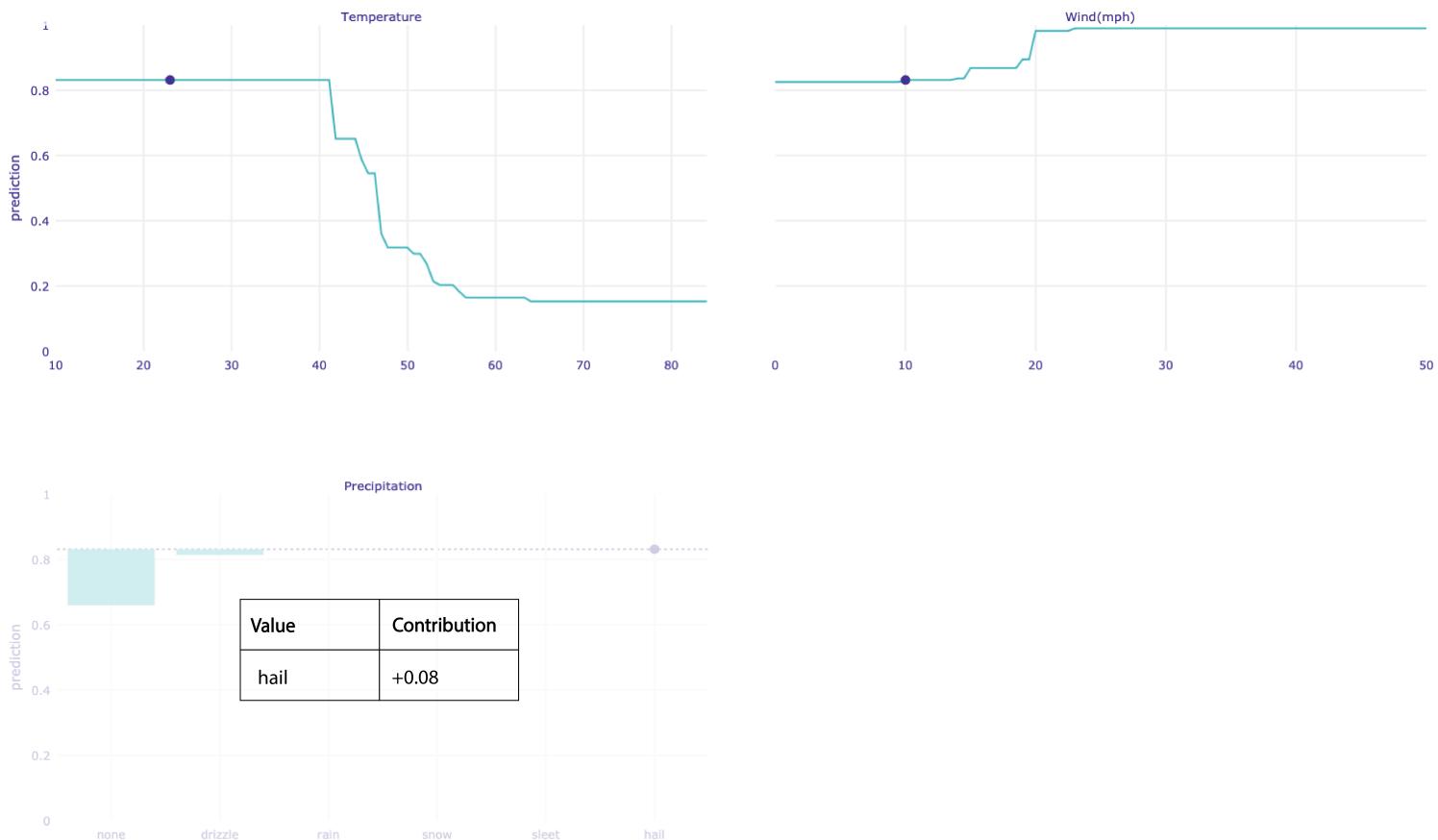
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861



Correct -

In this case, the **Contribution** value when you mouse over the Temperature chart has a positive (+) sign. Also, the area **below** the current prediction dot in the Temperature chart is large, while the area above the prediction dot is 0. Changing the value of the temperature is probably going to **decrease** the probability that you should wear a coat. So the current value of temperature is pushing the model toward predicting 'YES'.

The **Contribution** value when you mouse over the Precipitation chart has a positive (+) sign. Also, all of the bars for the Precipitation chart are pointing down **below** the blue dot. This means that changing precipitation is probably going to **decrease** the probability that you should wear a coat. So the current value of precipitation is also pushing the model toward predicting 'YES'.

Meanwhile, the **Contribution** value when you mouse over the Wind chart has a negative (-) sign. Also, the area **above** the current prediction dot in the Wind(mph) chart is medium size, while the area below the prediction dot is 0. Changing the value of wind(mph) is probably going to **increase** the probability that you should wear a coat. So the current value of wind is pushing the model toward predicting 'NO'.

Not quite -

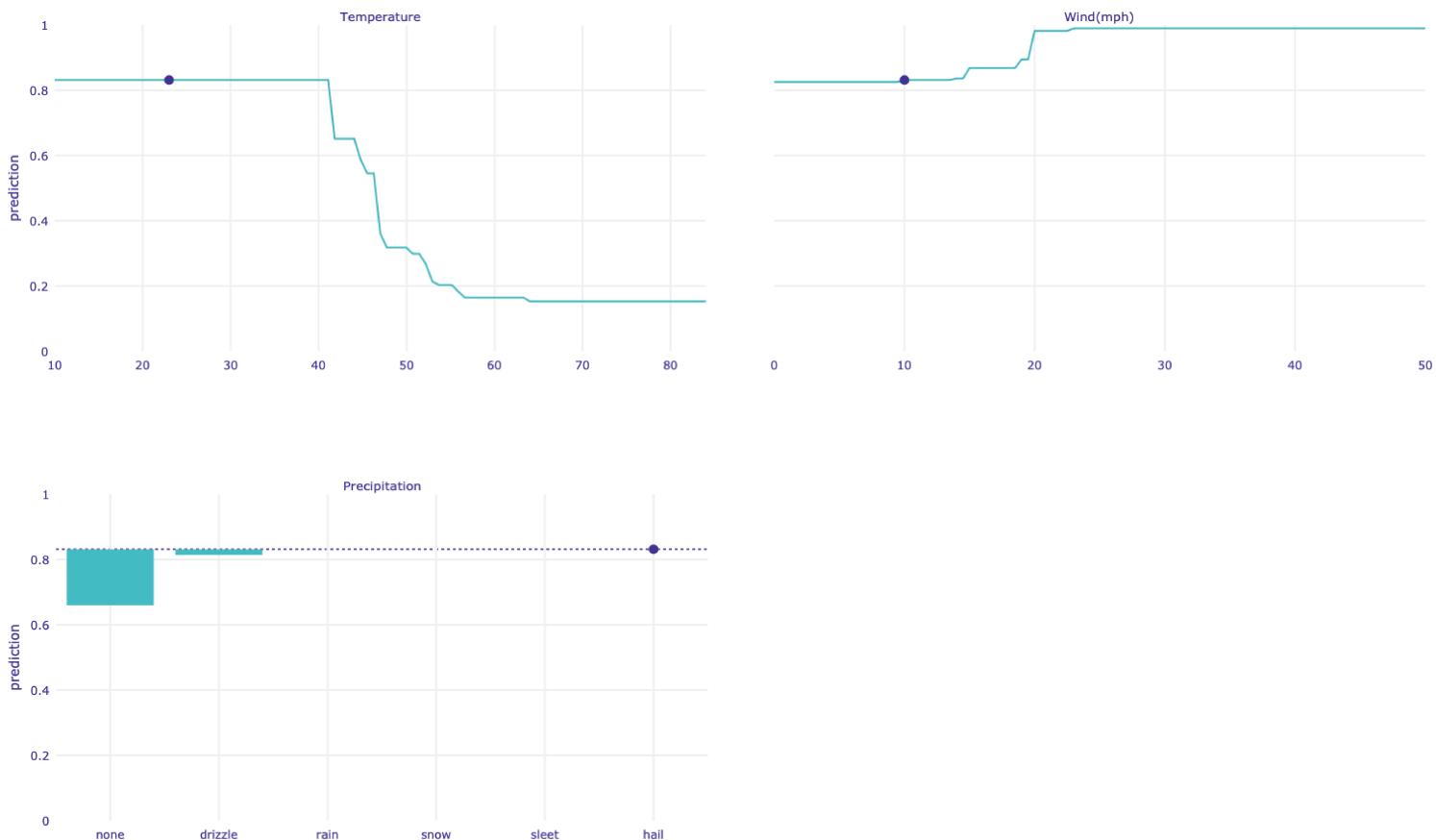
In this case, the **Contribution** value when you mouse over the Temperature chart has a positive (+) sign. Also, the area **below** the current prediction dot in the Temperature chart is large, while the area above the prediction dot is 0. Changing the value of the temperature is probably going to **decrease** the probability that

you should wear a coat. So the current value of temperature is pushing the model toward predicting 'YES'.

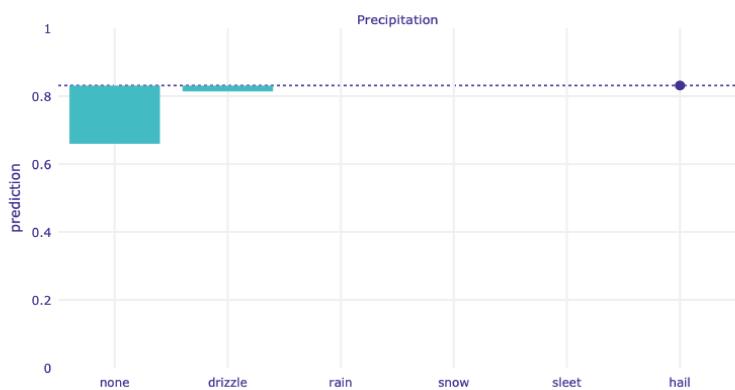
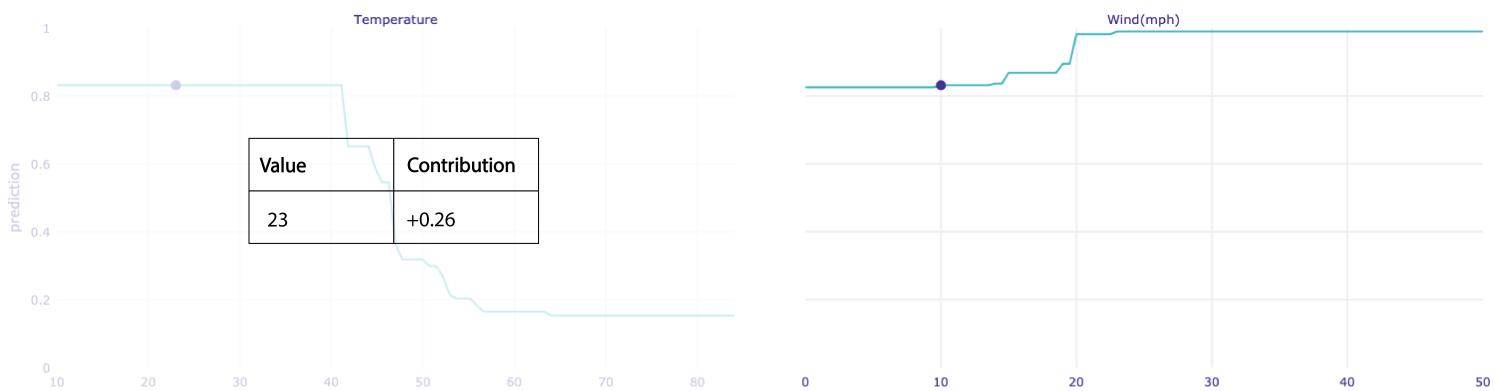
The **Contribution** value when you mouse over the Precipitation chart has a positive (+) sign. Also, all of the bars for the Precipitation chart are pointing down **below** the blue dot. This means that changing precipitation is probably going to **decrease** the probability that you should wear a coat. So the current value of precipitation is also pushing the model toward predicting 'YES'.

Meanwhile, the **Contribution** value when you mouse over the Wind chart has a negative (-) sign. Also, the area **above** the current prediction dot in the Wind(mph) chart is medium size, while the area below the prediction dot is 0. Changing the value of wind(mph) is probably going to **increase** the probability that you should wear a coat. So the current value of wind is pushing the model toward predicting 'NO'.

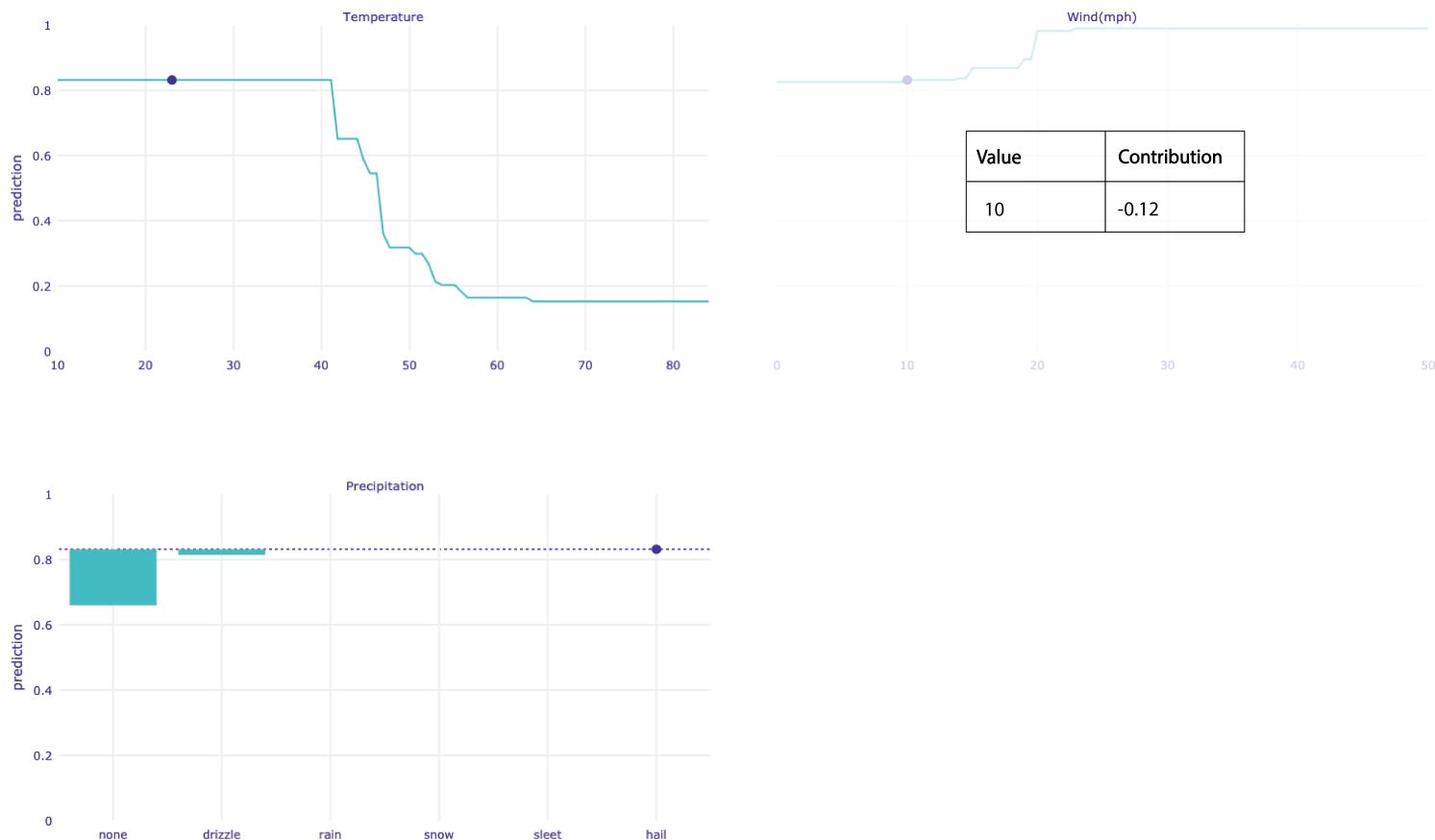
Model Prediction: 0.861

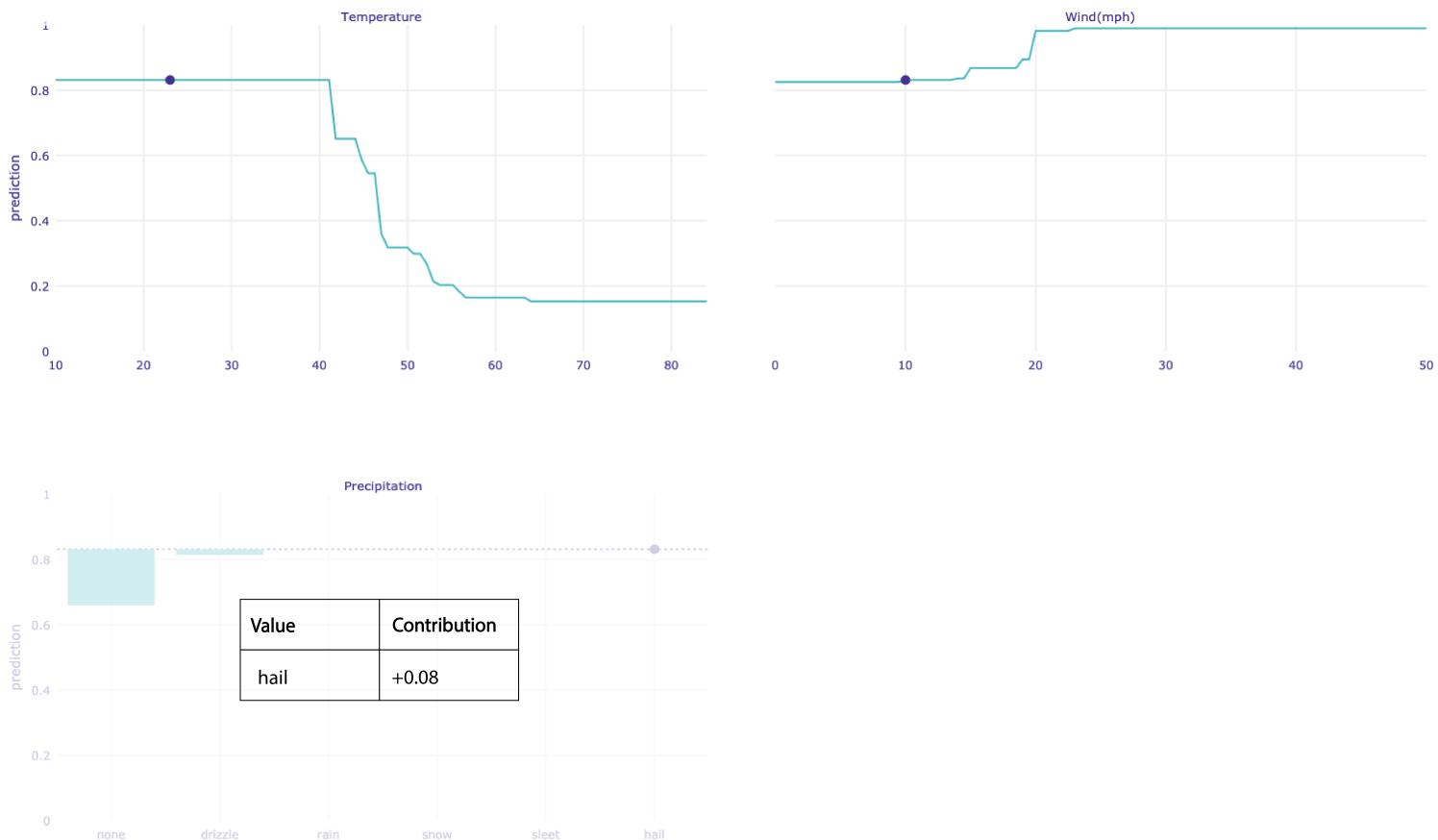


Model Prediction: 0.861



Model Prediction: 0.861

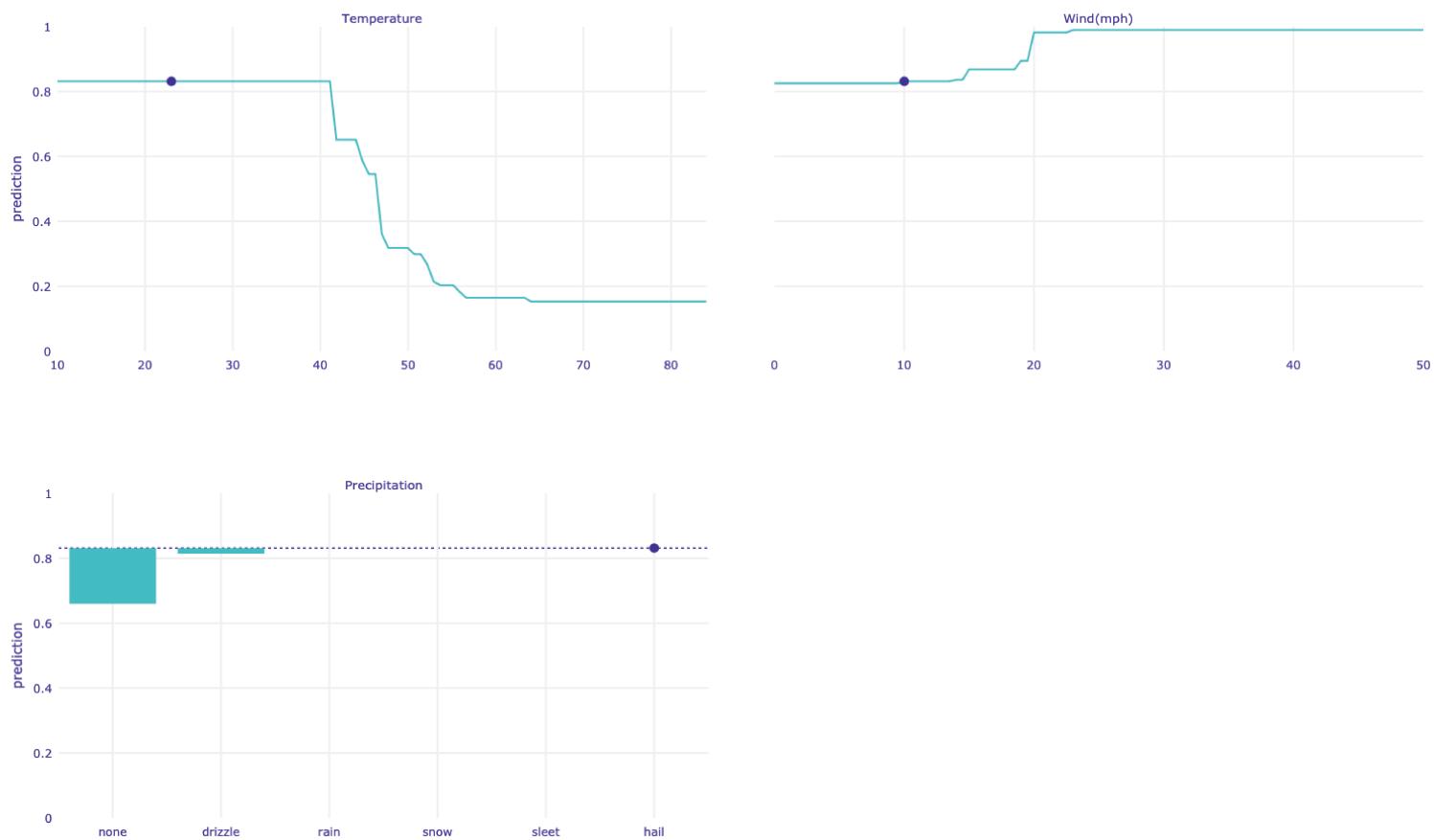


Model Prediction: 0.861

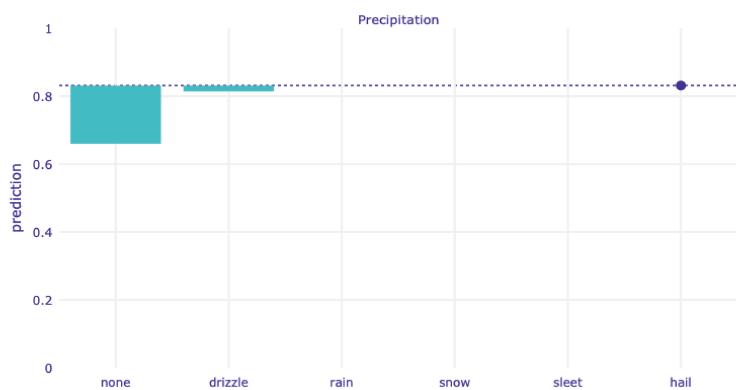
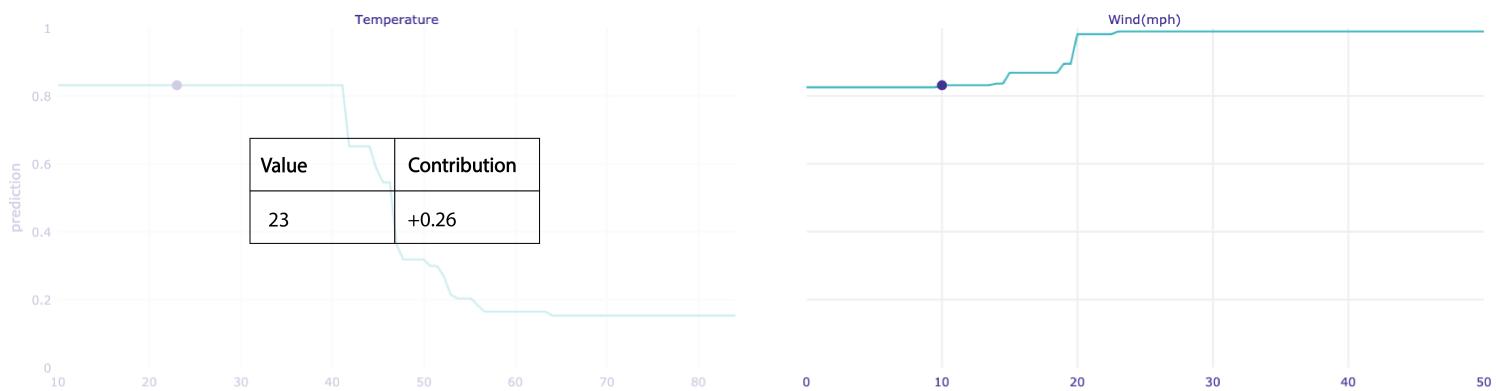
By looking at the explanation image, which factor(s) are pushing the model toward predicting 'NO'?

- Temperature
- Wind
- Precipitation

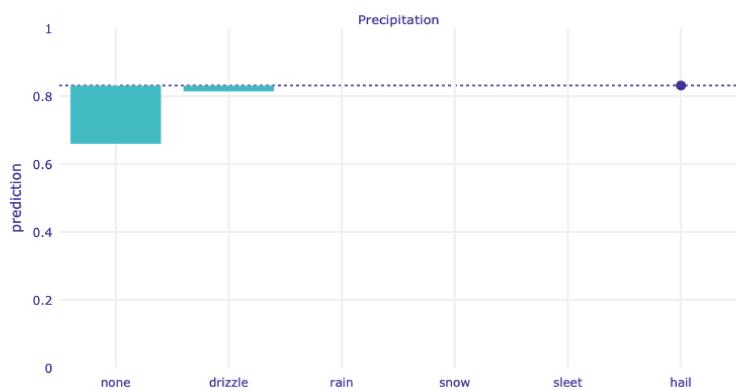
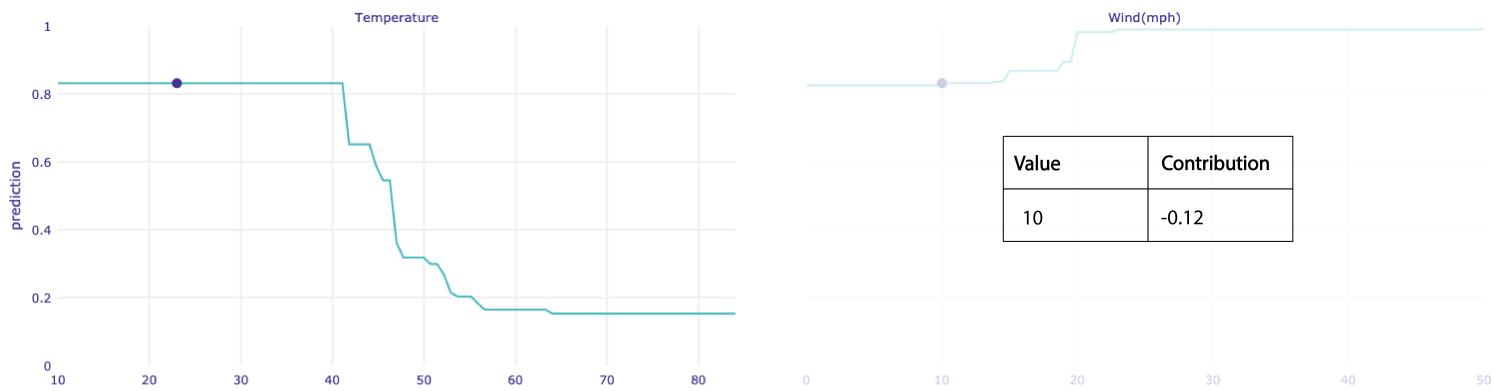
Model Prediction: 0.861



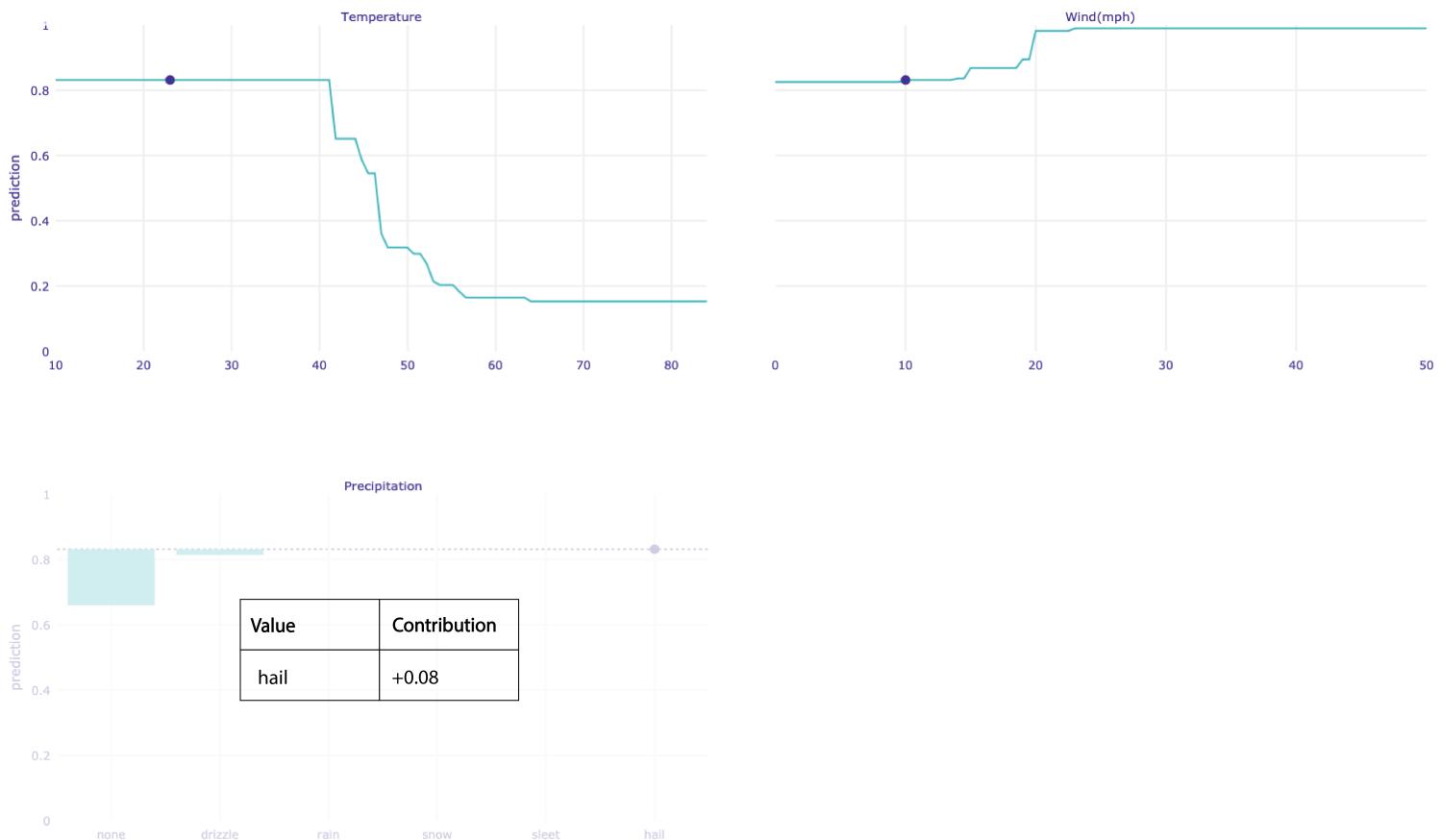
Model Prediction: 0.861



Model Prediction: 0.861



Model Prediction: 0.861



Correct -

In this case, the **Contribution** value when you mouse over the Temperature chart has a positive (+) sign. Also, the area **below** the current prediction dot in the Temperature chart is large, while the area above the prediction dot is 0. Changing the value of the temperature is probably going to **decrease** the probability that you should wear a coat. So the current value of temperature is pushing the model toward predicting 'YES'.

The **Contribution** value when you mouse over the Precipitation chart has a positive (+) sign. Also, all of the bars for the Precipitation chart are pointing down **below** the blue dot. This means that changing precipitation is probably going to **decrease** the probability that you should wear a coat. So the current value of precipitation is also pushing the model toward predicting 'YES'.

Meanwhile, the **Contribution** value when you mouse over the Wind chart has a negative (-) sign. Also, the area **above** the current prediction dot in the Wind(mph) chart is medium size, while the area below the prediction dot is 0. Changing the value of wind(mph) is probably going to **increase** the probability that you should wear a coat. So the current value of wind is pushing the model toward predicting 'NO'.

Not quite -

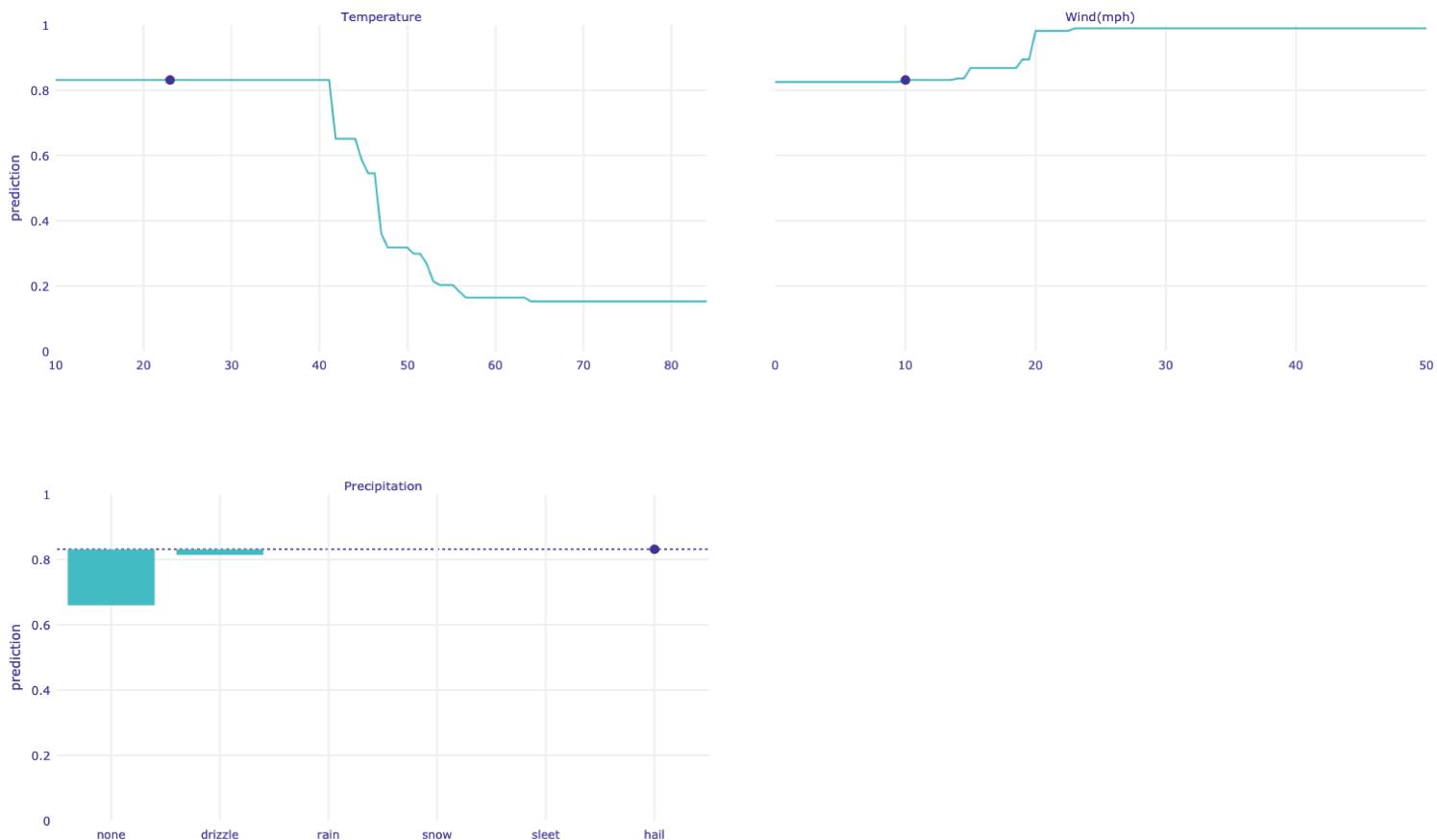
In this case, the **Contribution** value when you mouse over the Temperature graph has a positive (+) sign. Also, the area **below** the current prediction dot in the Temperature chart is large, while the area above the prediction dot is 0. Changing the value of the temperature is probably going to **decrease** the probability that

you should wear a coat. So the current value of temperature is pushing the model toward predicting 'YES'.

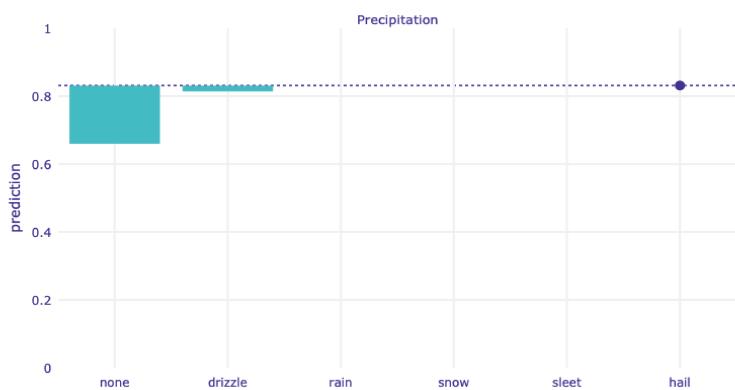
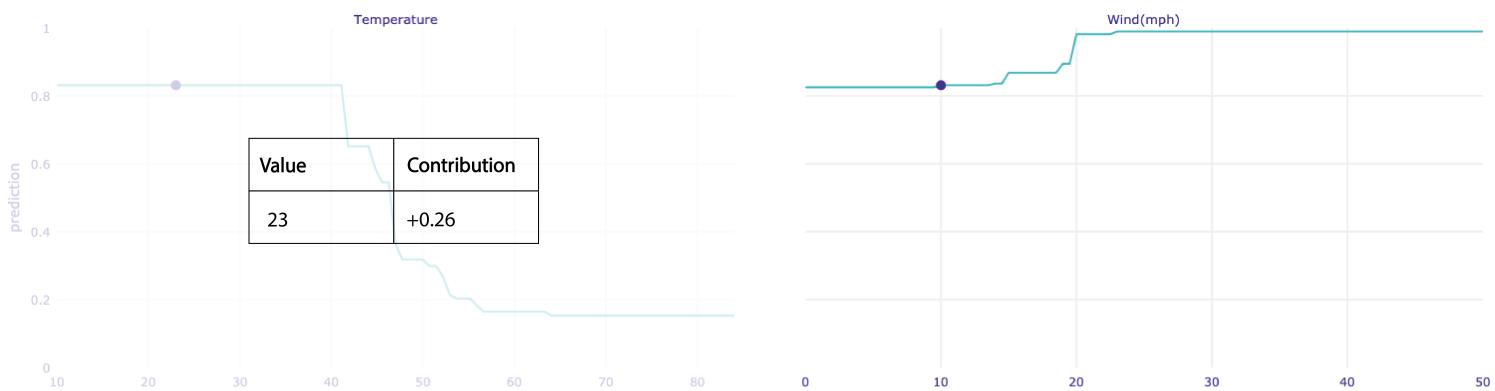
The **Contribution** value when you mouse over the Precipitation graph has a positive (+) sign. Also, all of the bars for the Precipitation chart are pointing down **below** the blue dot. This means that changing precipitation is probably going to **decrease** the probability that you should wear a coat. So the current value of precipitation is also pushing the model toward predicting 'YES'.

Meanwhile, the **Contribution** value when you mouse over the Wind graph has a negative (-) sign. Also, the area **above** the current prediction dot in the Wind(mph) chart is medium size, while the area below the prediction dot is 0. Changing the value of wind(mph) is probably going to **increase** the probability that you should wear a coat. So the current value of wind is pushing the model toward predicting 'NO'.

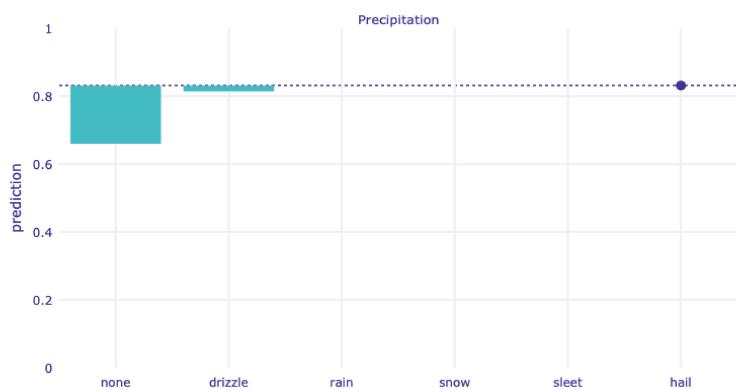
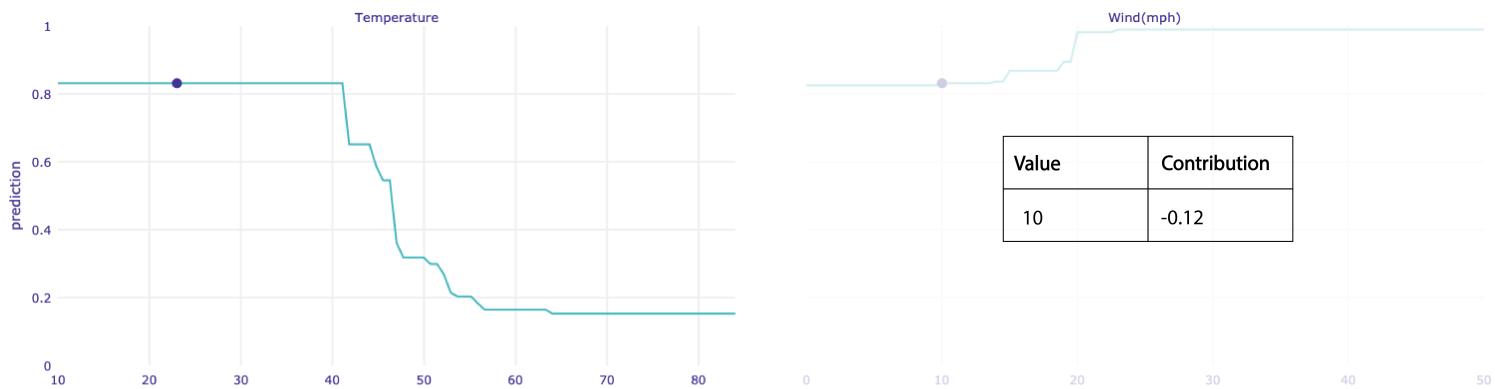
Model Prediction: 0.861

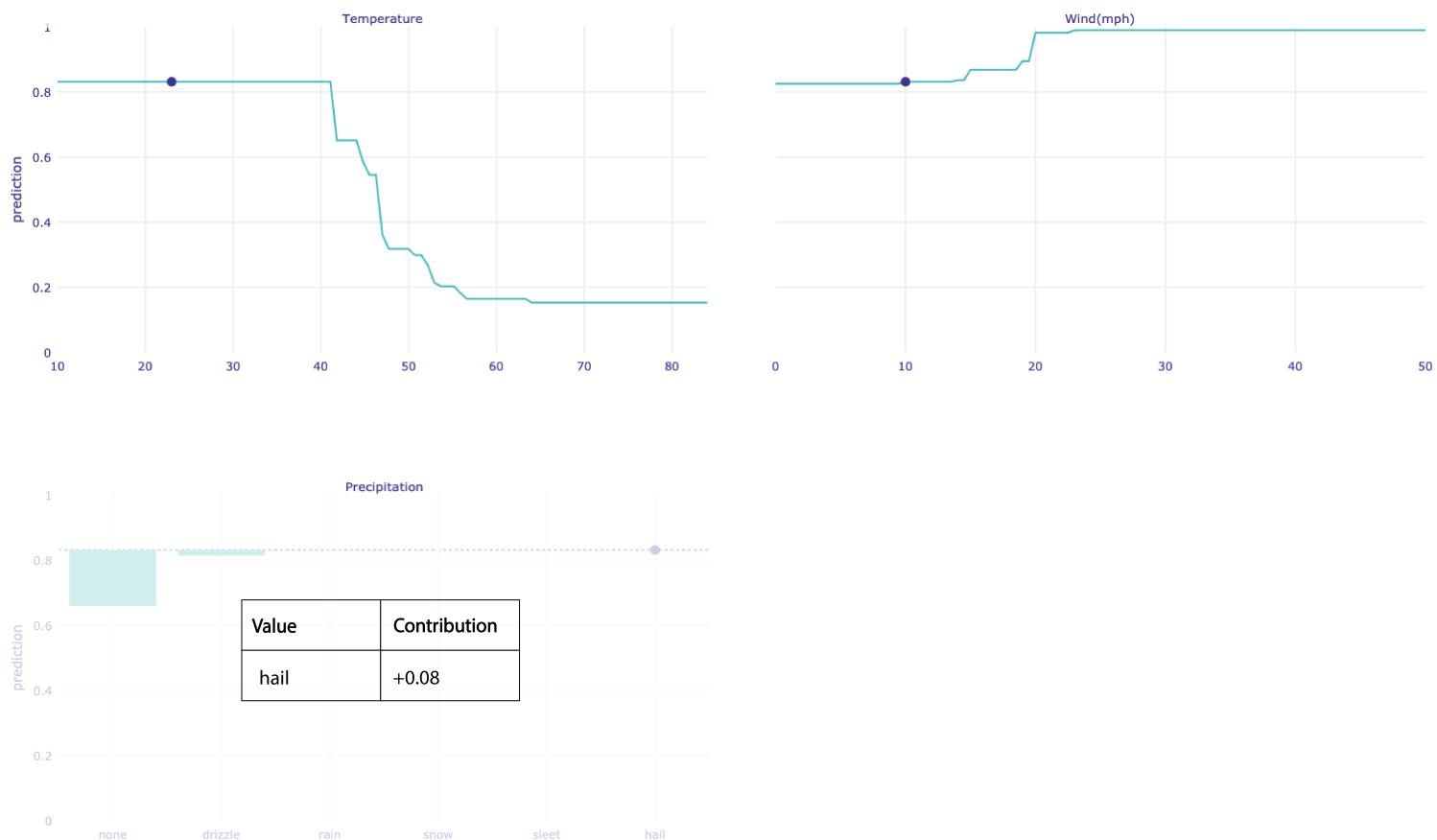


Model Prediction: 0.861



Model Prediction: 0.861

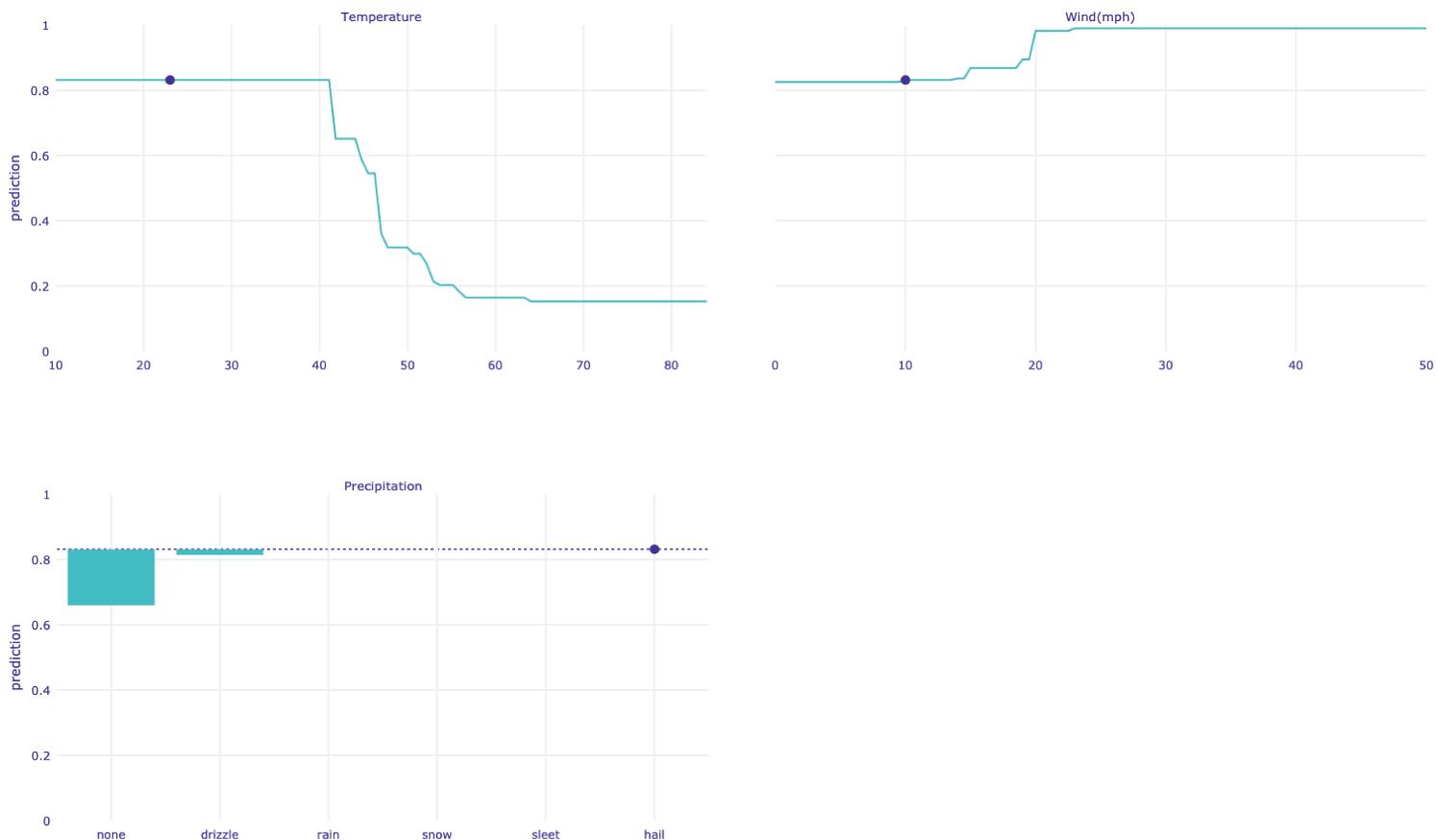


Model Prediction: 0.861

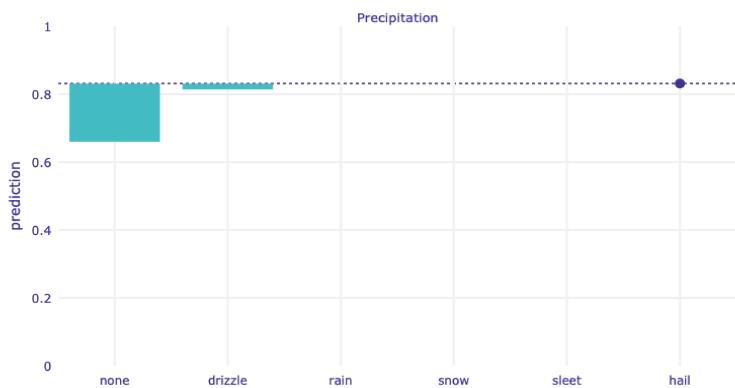
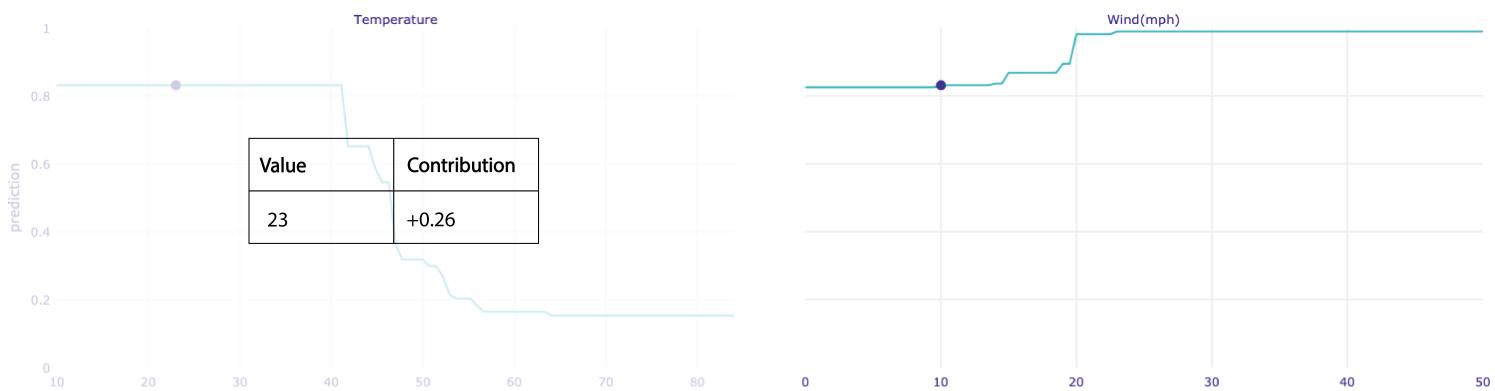
Which factor has the greatest predictive power?

- Temperature
- Wind
- Precipitation

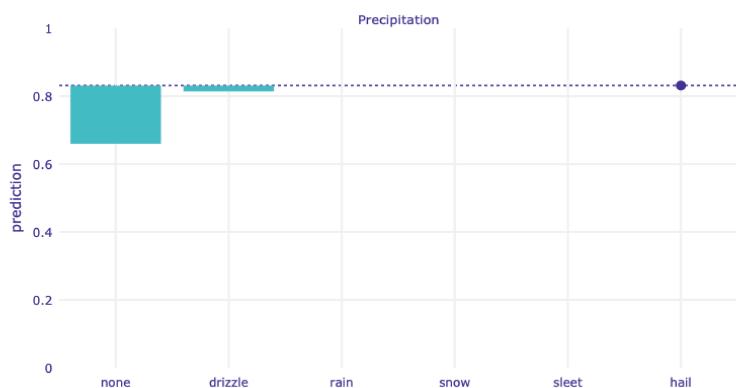
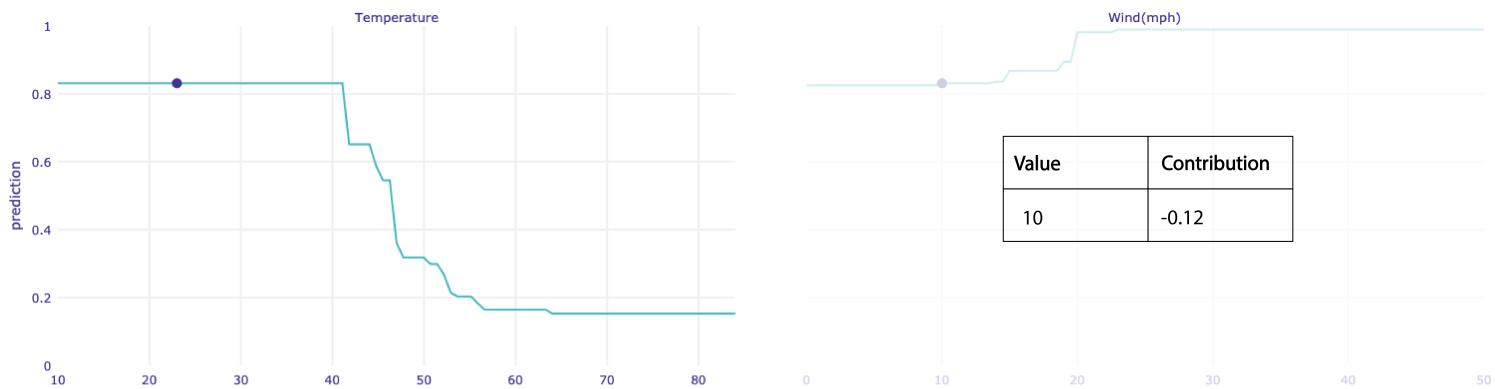
Model Prediction: 0.861



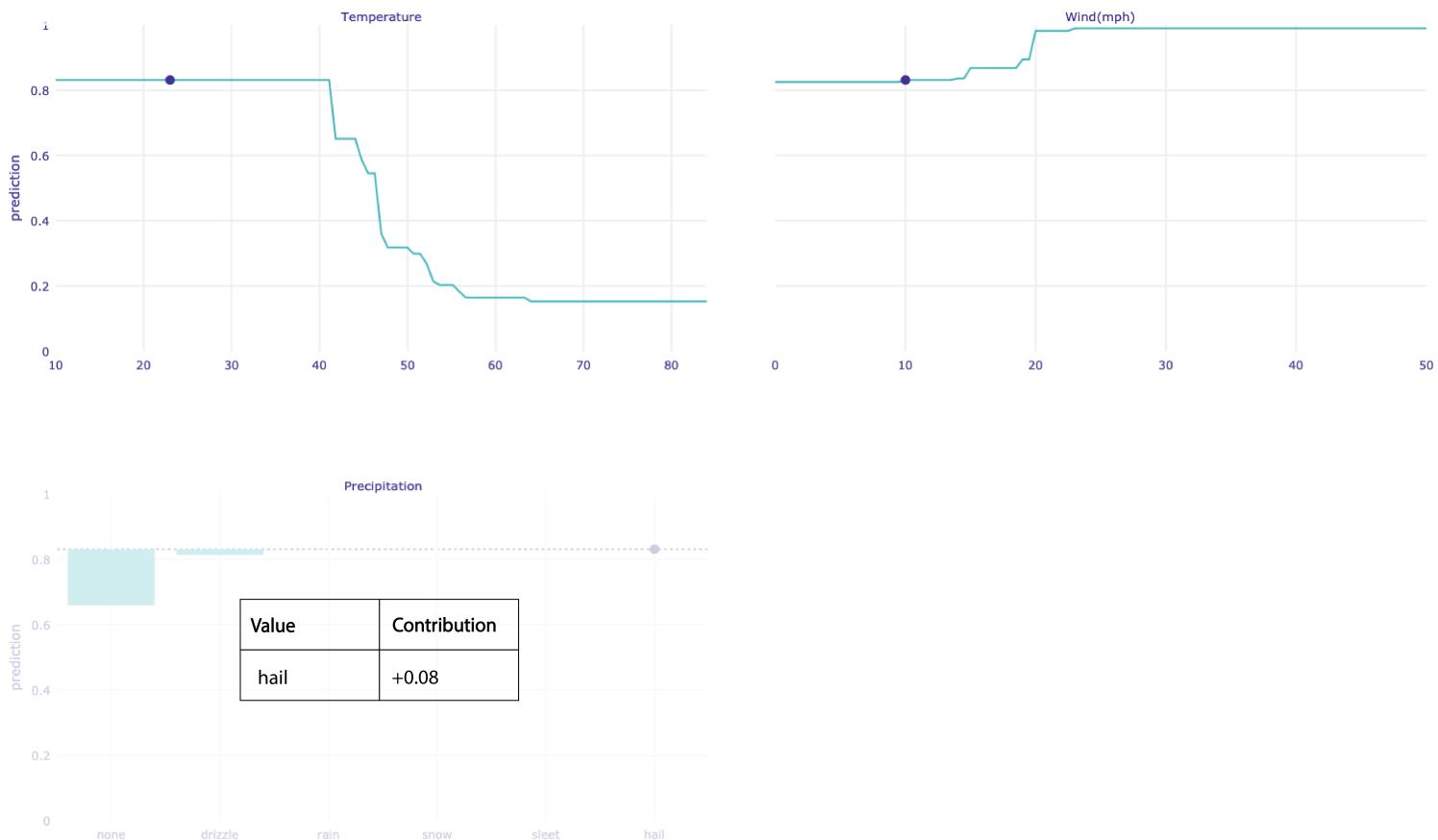
Model Prediction: 0.861



Model Prediction: 0.861



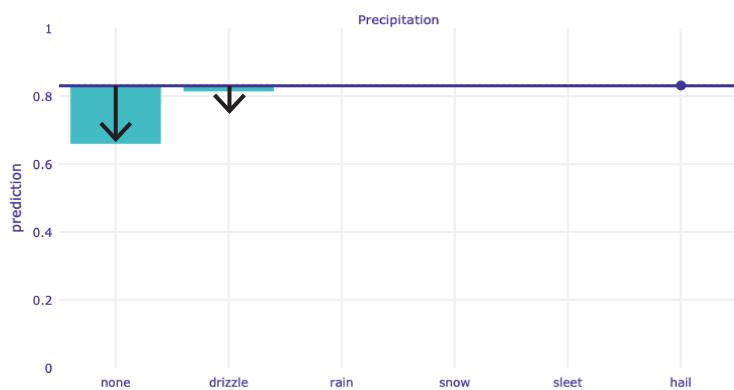
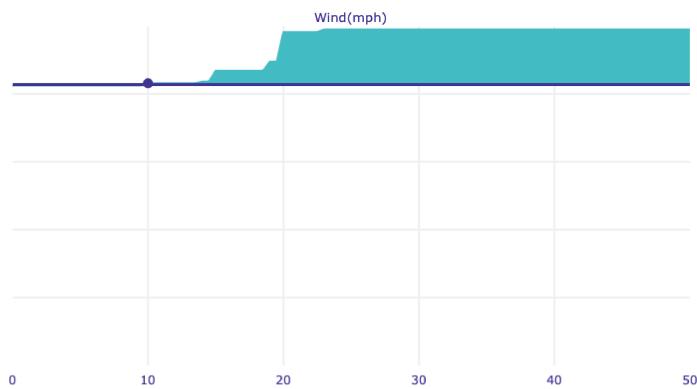
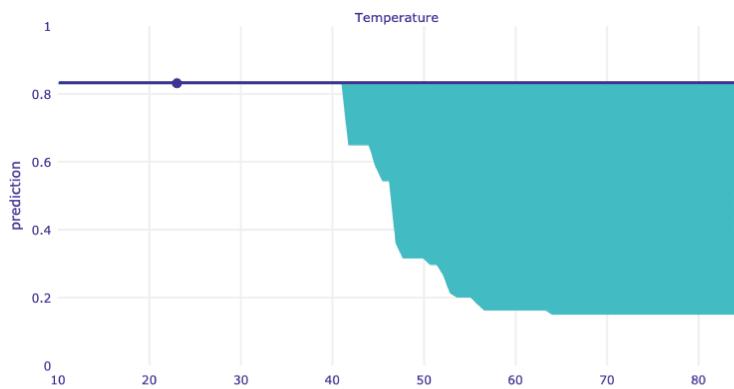
Model Prediction: 0.861



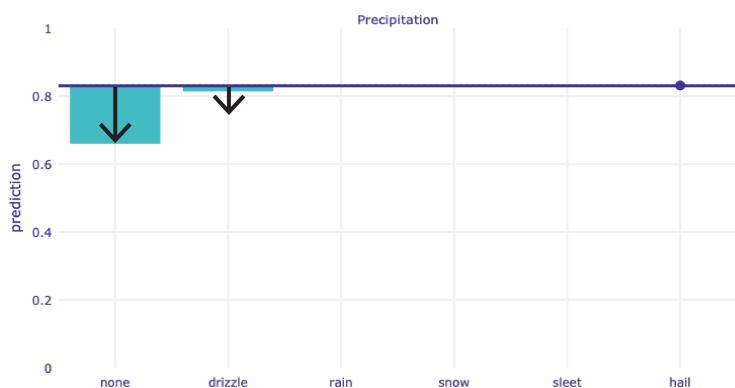
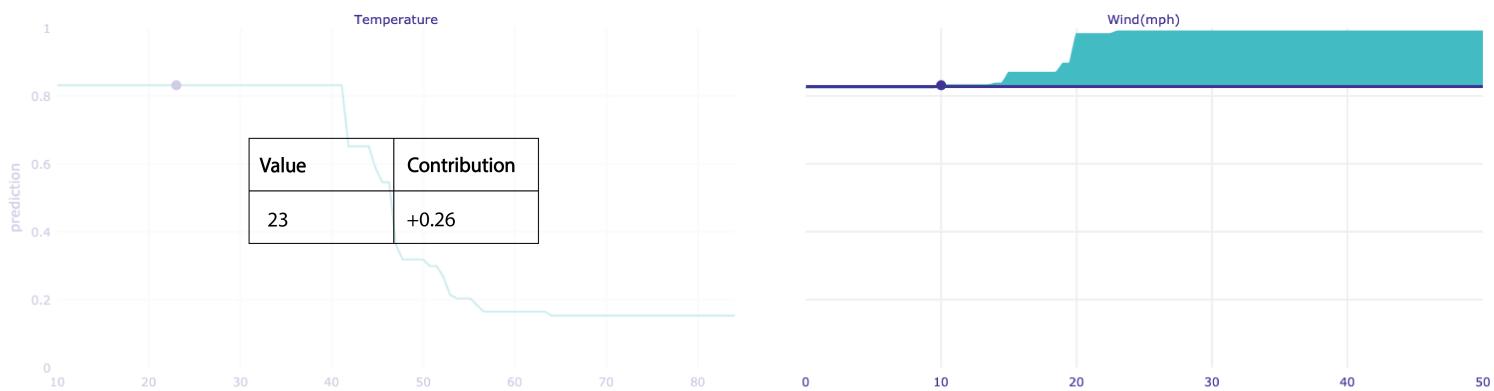
Correct – Temperature has the largest predictive power, because its contribution value of 0.26 is larger than 0.12 (Wind) or 0.08 (Precipitation). Also, the area between the line and the current prediction in the Temperature chart is larger than the area between the line and the current prediction in the Wind(mph) chart, and the bars in the Precipitation chart. Changing the Temperature can cause a larger change in model output than changing the Wind(mph) or the Precipitation.

Not quite - Temperature has the largest predictive power, because its contribution value of 0.26 is larger than 0.12 (Wind) or 0.08 (Precipitation). Also, the area between the line and the current prediction in the Temperature chart is larger than the area between the line and the current prediction in the Wind(mph) chart, and the bars in the Precipitation chart. Changing the Temperature can cause a larger change in model output than changing the Wind(mph) or the Precipitation.

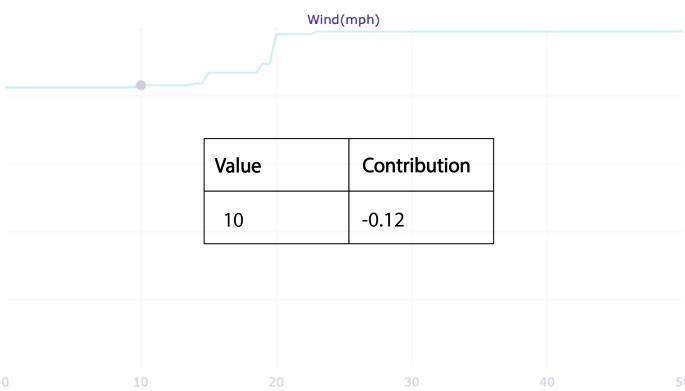
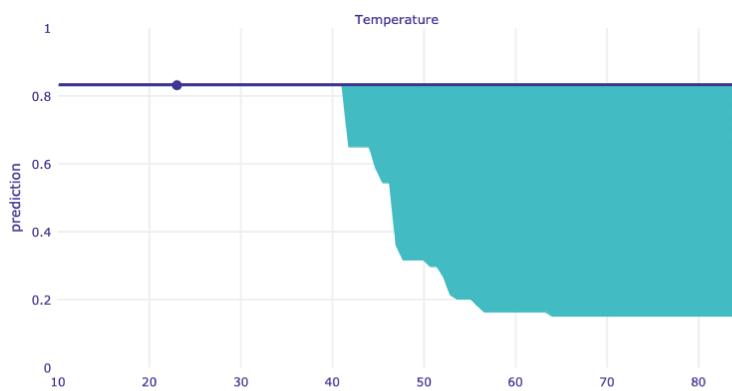
Model Prediction: 0.861



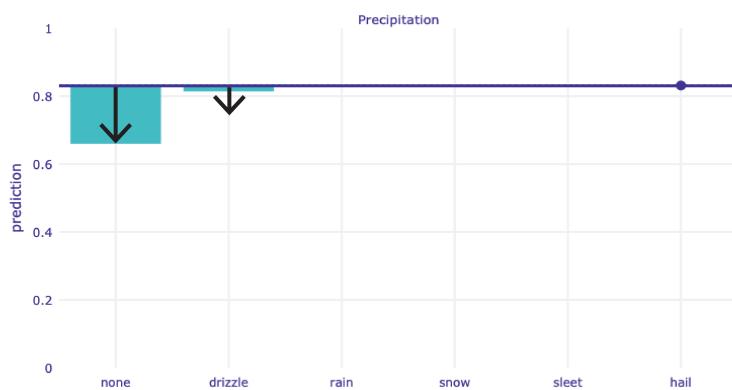
Model Prediction: 0.861



Model Prediction: 0.861



Value	Contribution
10	-0.12



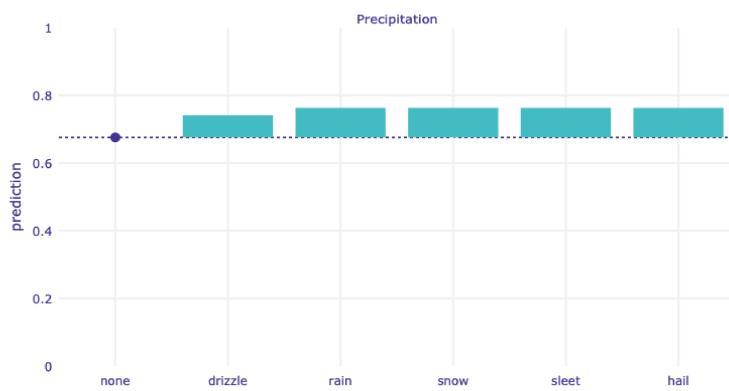
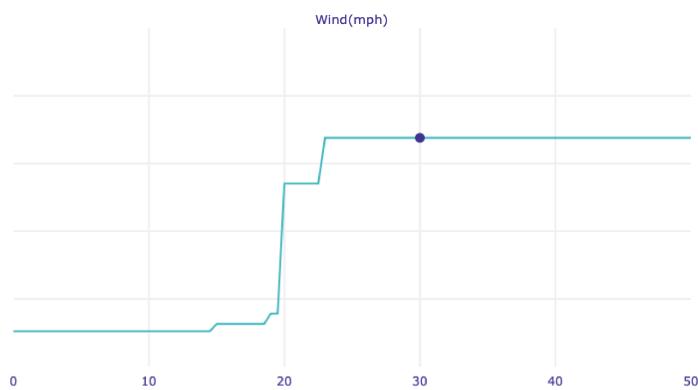
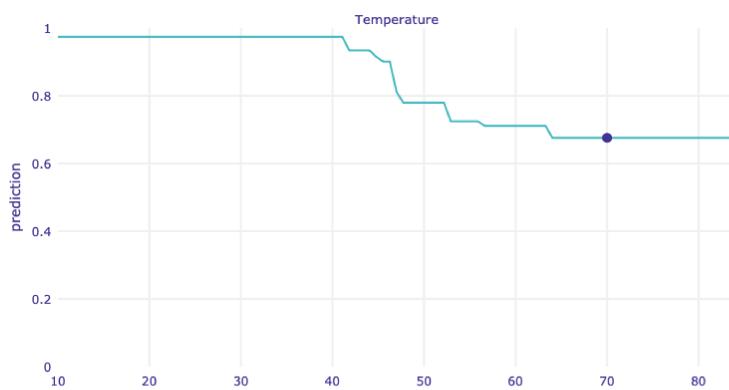
Model Prediction: 0.861

Intro Test 4

As a final review, what does the following model recommend you do?

- YES, you should wear a coat
- NO, do not wear a coat

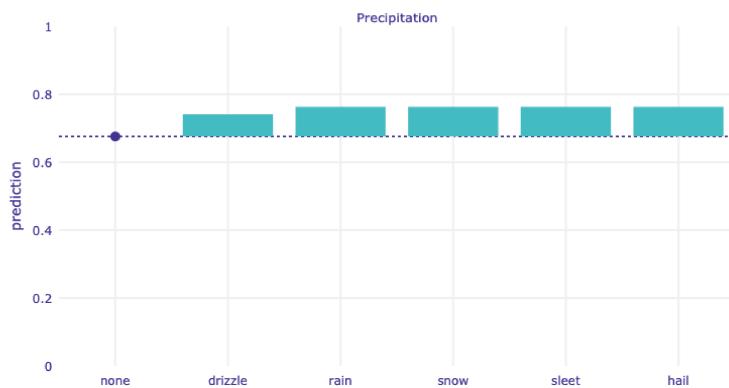
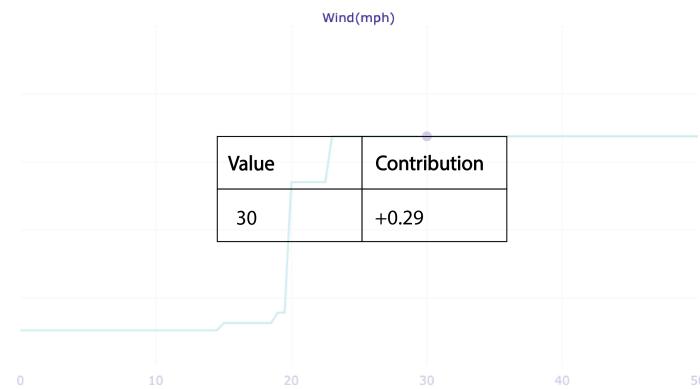
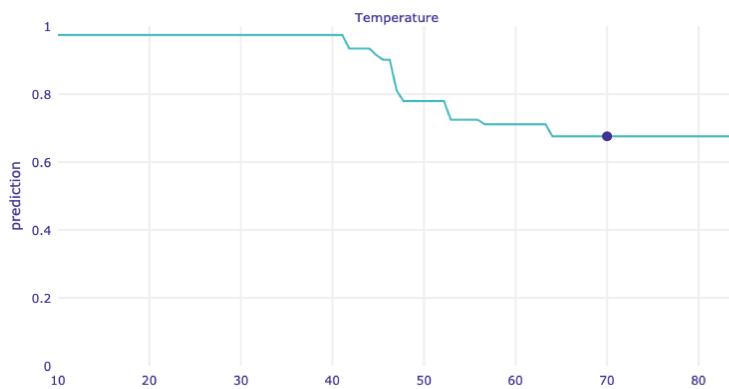
Model Prediction: 0.648

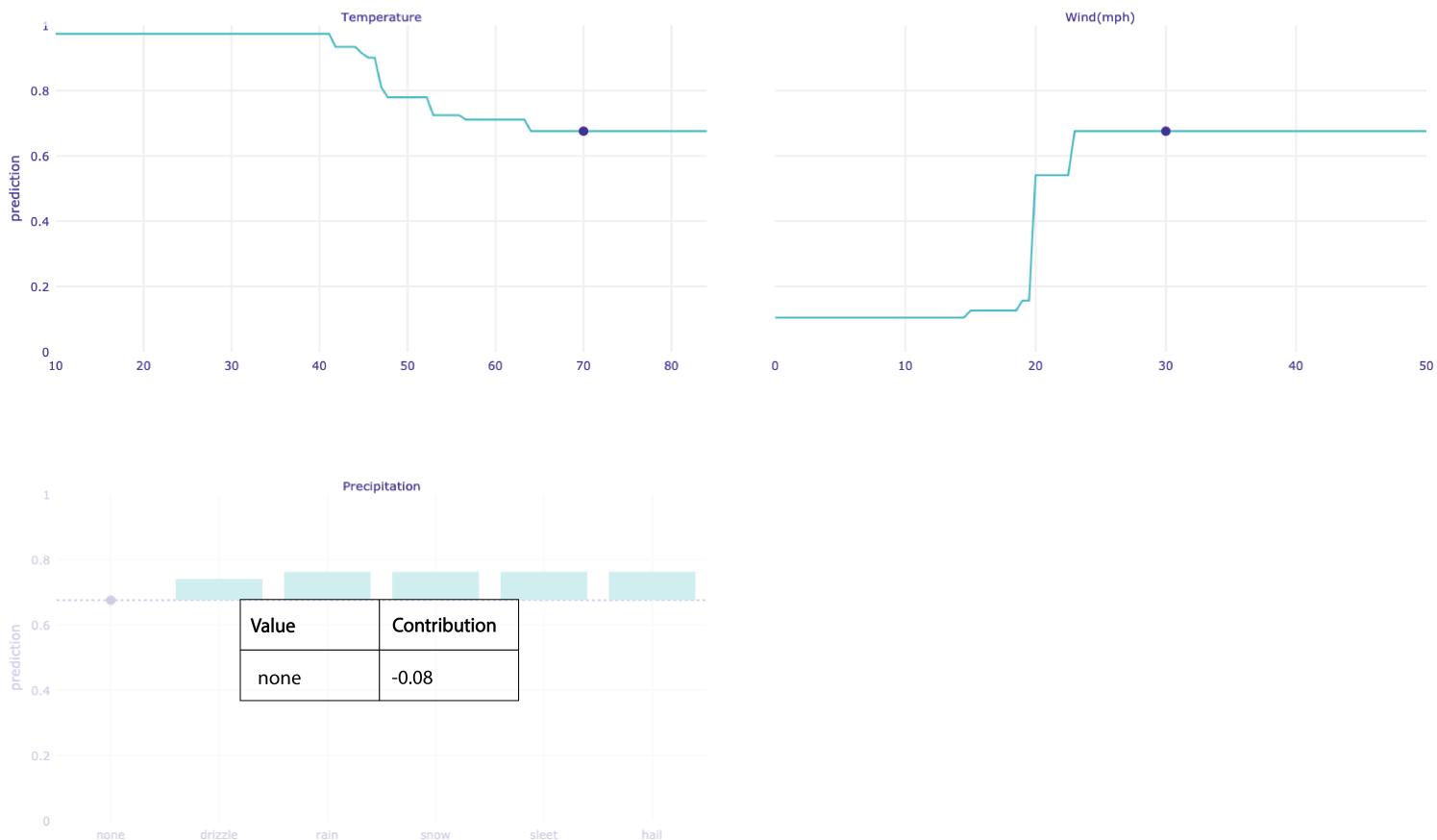


Model Prediction: 0.648



Model Prediction: 0.648

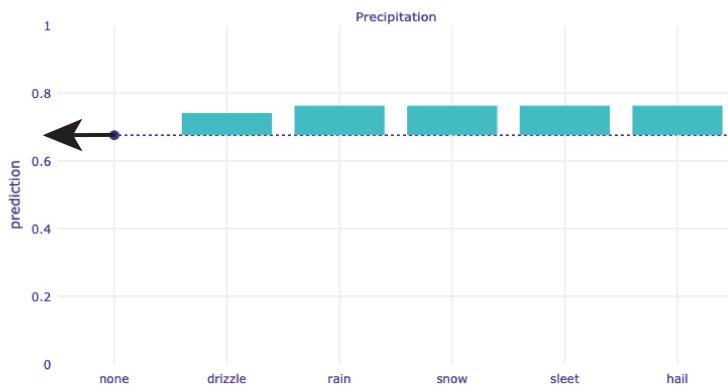
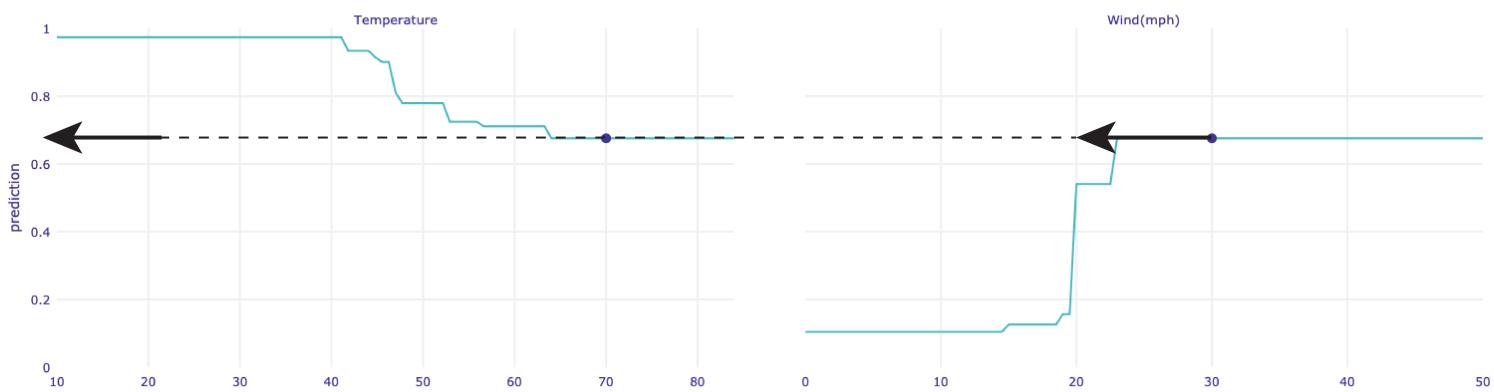


Model Prediction: 0.648

Correct. In this case, the model prediction is 0.648, which is greater than 0.5, so the model will return 'YES'.

Incorrect. In this case, the model prediction is 0.648, which is greater than 0.5, so the model will return 'YES'.

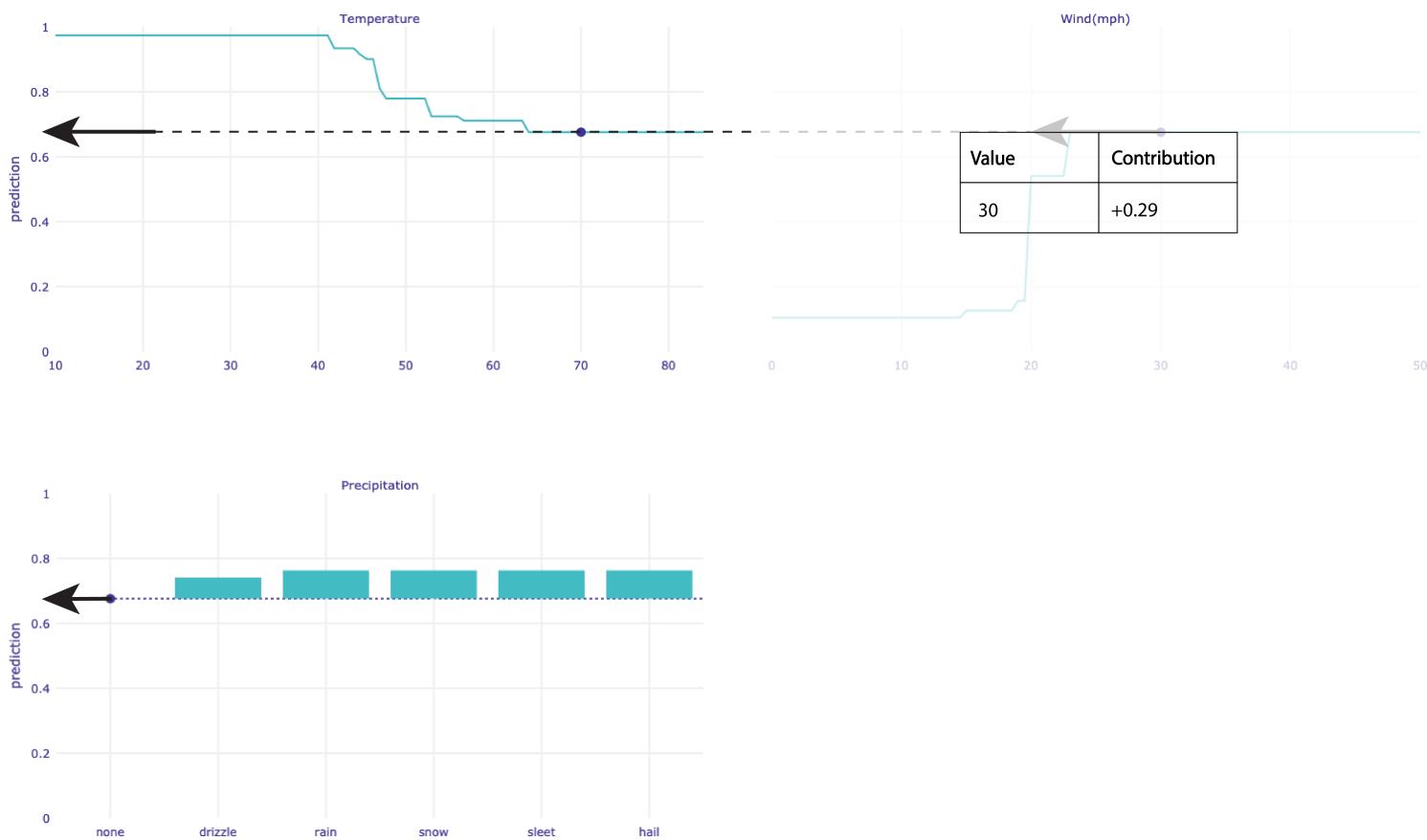
Model Prediction: 0.648



Model Prediction: 0.648



Model Prediction: 0.648



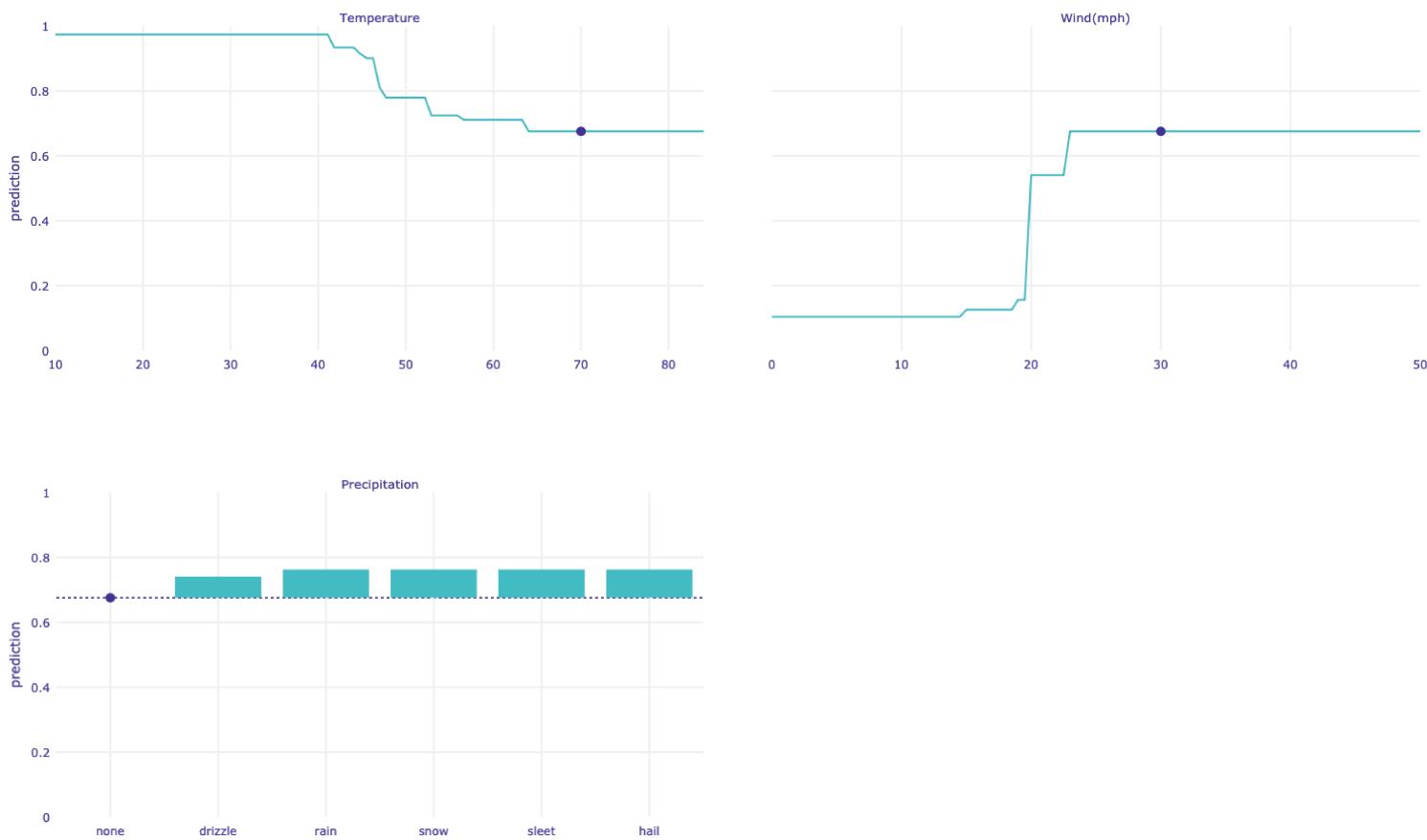
Model Prediction: 0.648



By looking at the explanation image, please select the value for **temperature** input into the model:

- 84
- 70
- 61
- 56
- 37

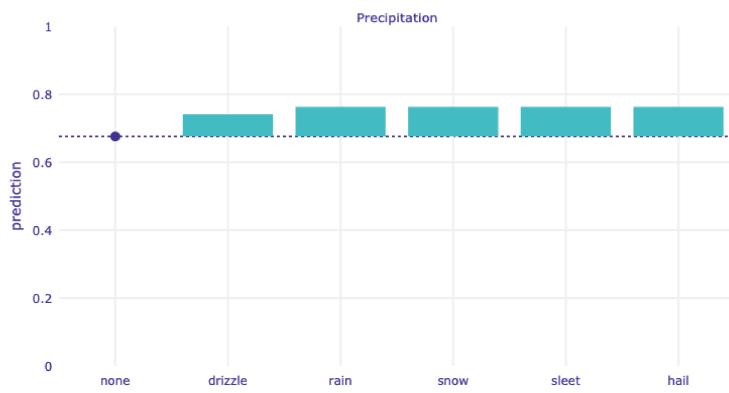
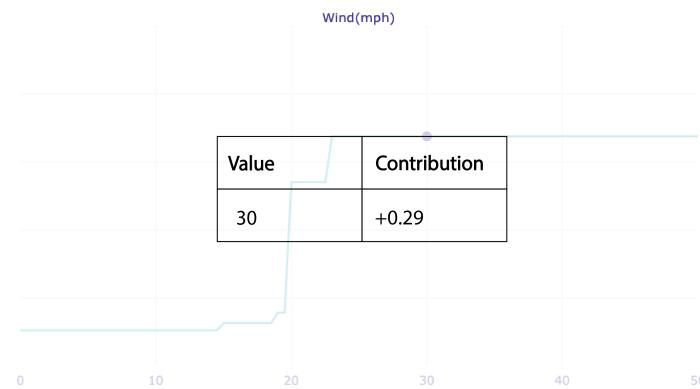
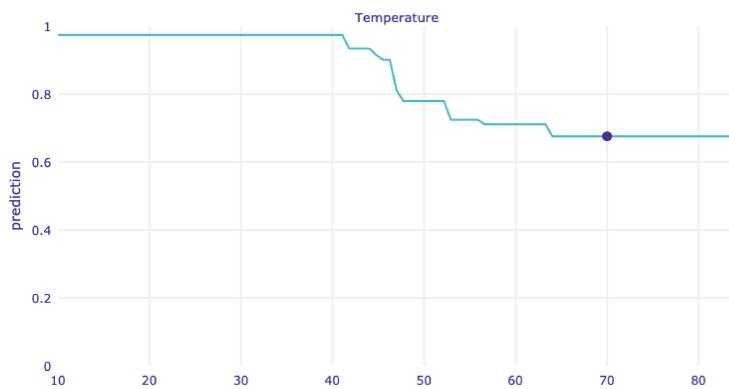
Model Prediction: 0.648



Model Prediction: 0.648



Model Prediction: 0.648



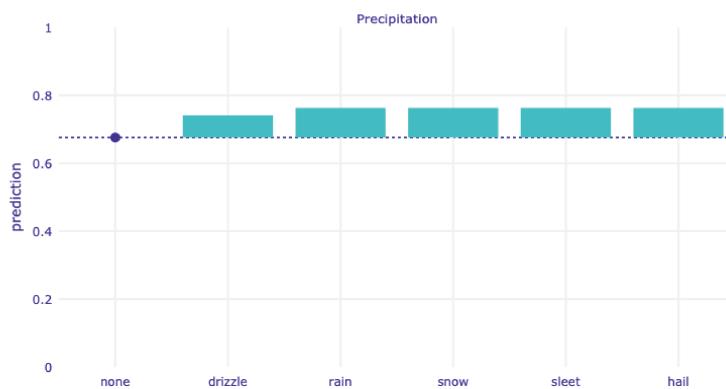
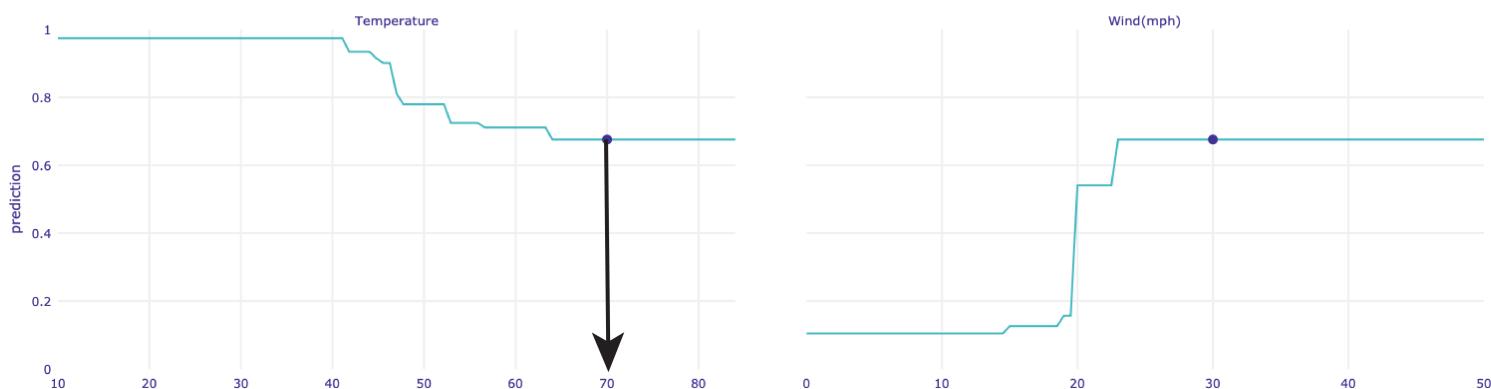
Model Prediction: 0.648

Correct – The value is in the **Value** column when you mouse over the Temperature chart. It is also under the **blue dot** in the Temperature chart. This value is 70.

Incorrect – The value is in the **Value** column when you mouse

over the Temperature chart. It is also under the **blue dot** in the Temperature chart. This value is 70.

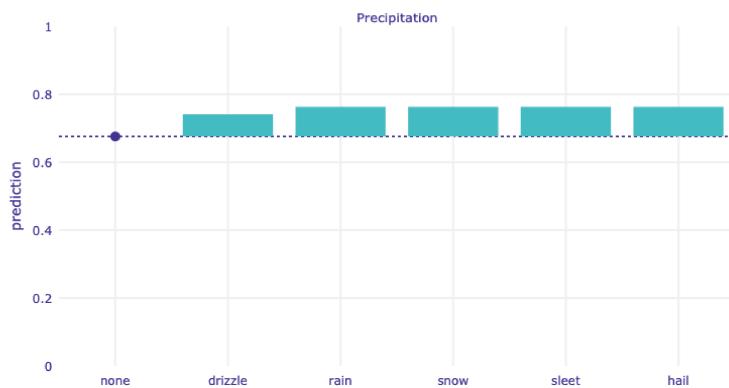
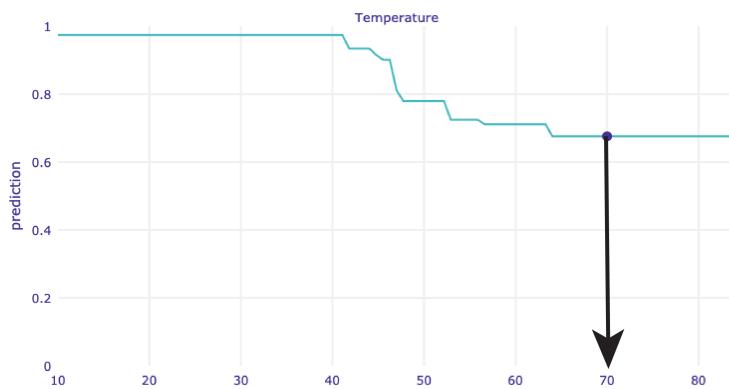
Model Prediction: 0.648



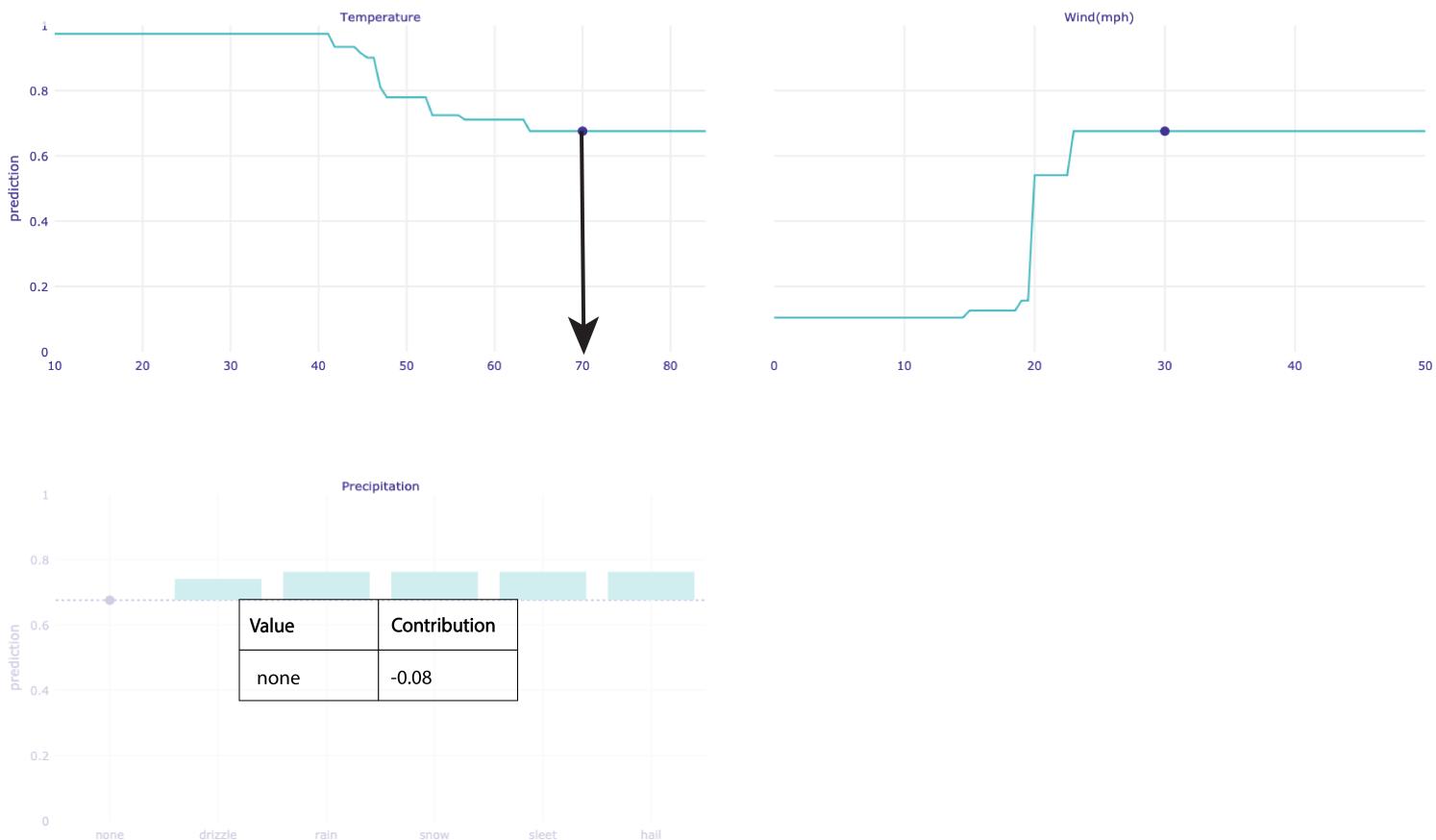
Model Prediction: 0.648



Model Prediction: 0.648



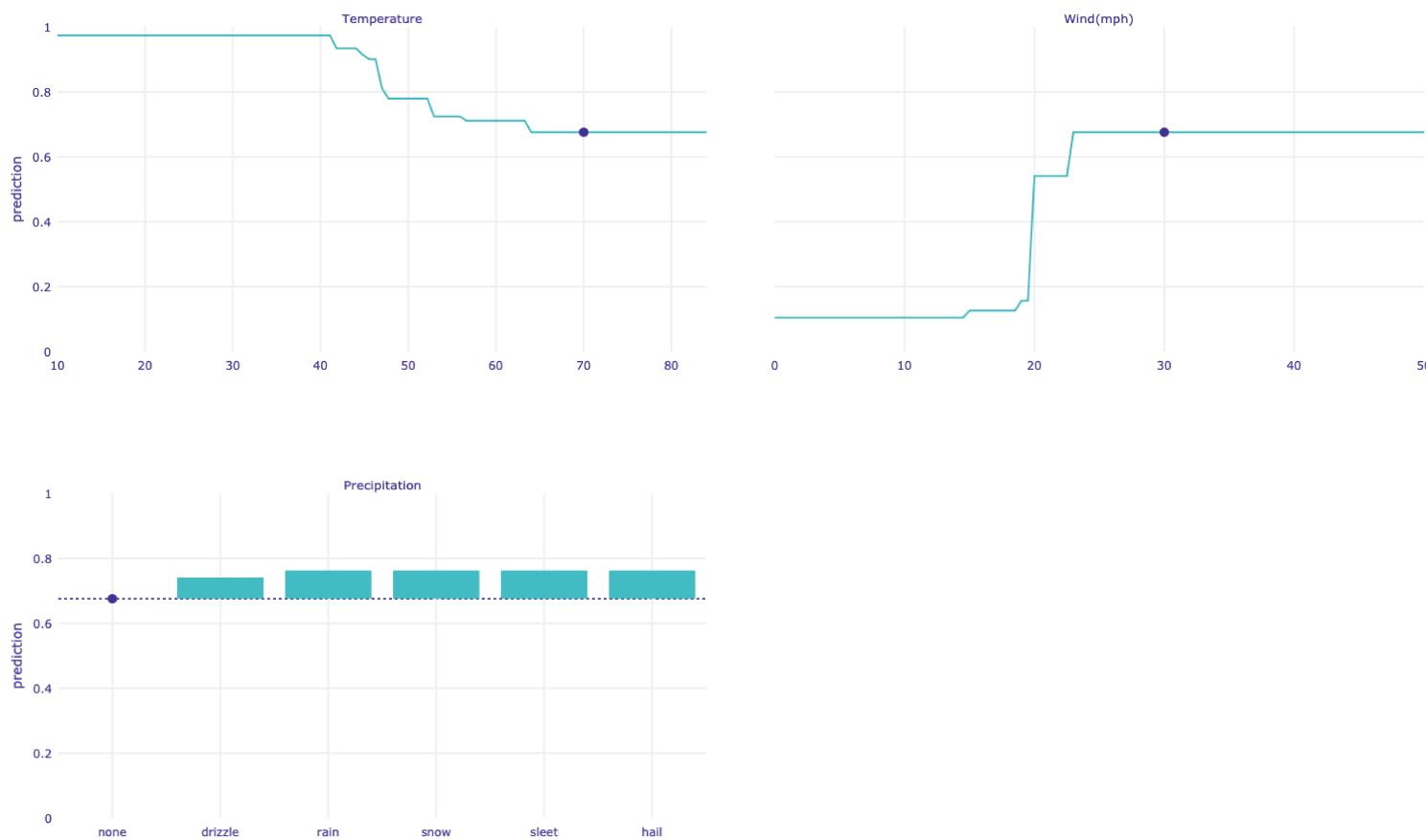
Model Prediction: 0.648



By looking at the explanation image, which factor(s) are pushing the model toward predicting 'NO?

- Temperature
- Wind
- Precipitation

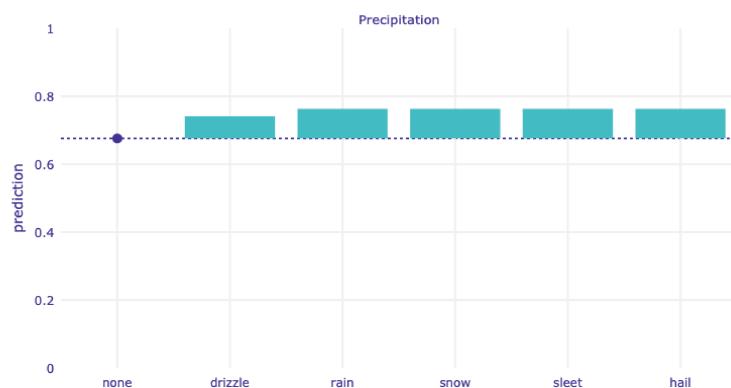
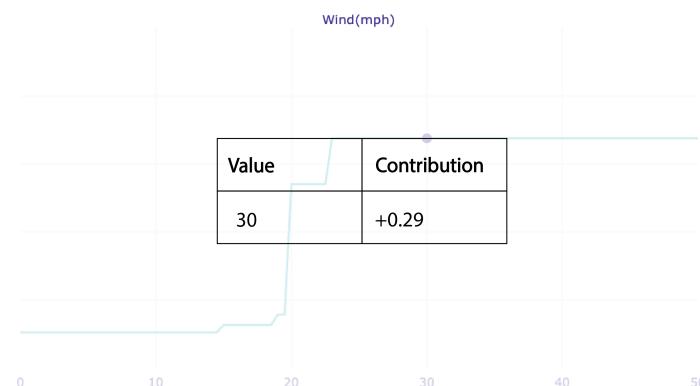
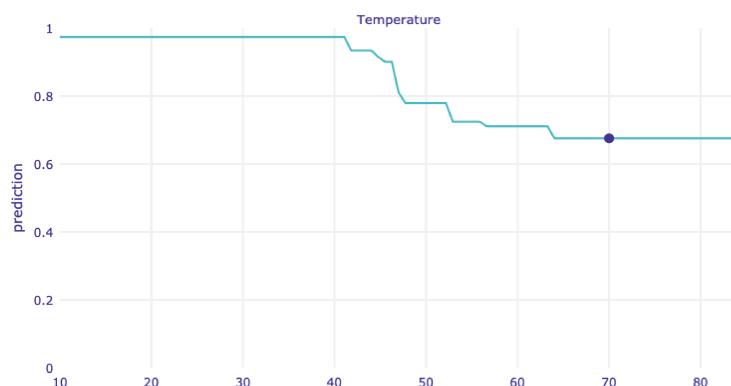
Model Prediction: 0.648



Model Prediction: 0.648



Model Prediction: 0.648



Model Prediction: 0.648



Correct -

The **Contribution** value when you mouse over the Temperature chart has a **negative (-)** sign. Also, the area **above** the current prediction dot in the Temperature chart is large, while the area below the prediction dot is 0. Changing the value of the temperature is probably going to **increase** the probability that you should wear a coat. So the current value of temperature is pushing the model toward predicting 'NO'.

The **Contribution** value when you mouse over the Precipitation chart has a **negative (-)** sign. Also, all of the bars for the Precipitation chart are pointing up **above** the prediction dot. This means that changing precipitation is probably going to **increase** the probability that you should wear a coat. So the current value of precipitation is also pushing the model toward predicting 'NO'.

Meanwhile, the **Contribution** value when you mouse over the Wind(mph) chart has a **positive (+)** sign. the area **below** the current prediction line in the Wind(mph) chart is large, while the area above the prediction line is 0. Changing the value of wind(mph) is probably going to **decrease** the probability that you should wear a coat. So the current value of wind is pushing the model toward predicting 'YES'.

Not quite -

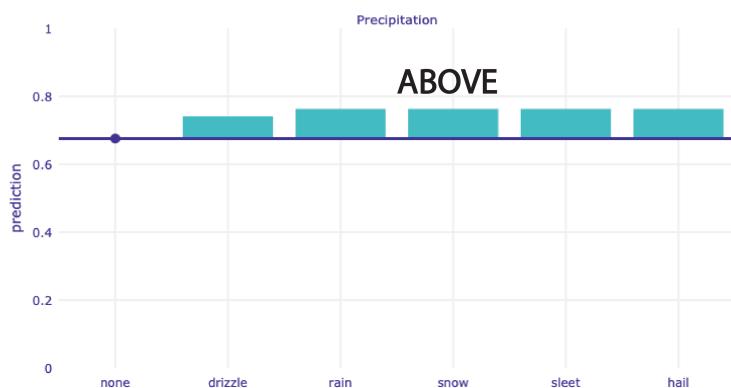
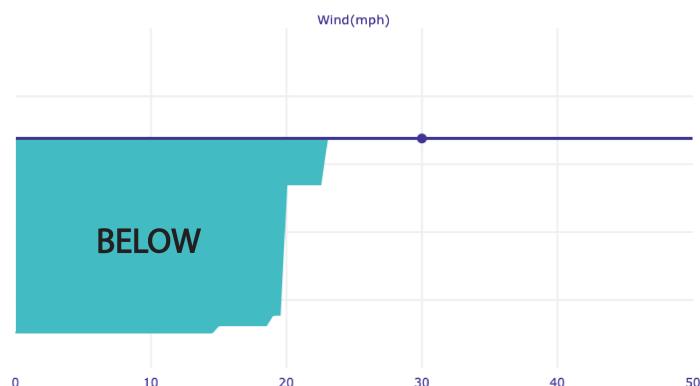
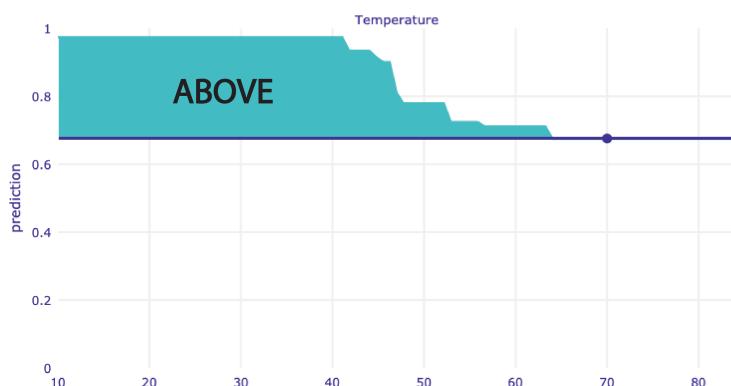
The **Contribution** value when you mouse over the Temperature chart has a **negative (-)** sign. Also, the area **above** the current prediction dot in the Temperature chart is large, while the area below the prediction dot is 0. Changing the value of the temperature is probably going to **increase** the probability that you should wear a coat. So the current value of temperature is

pushing the model toward predicting 'NO'.

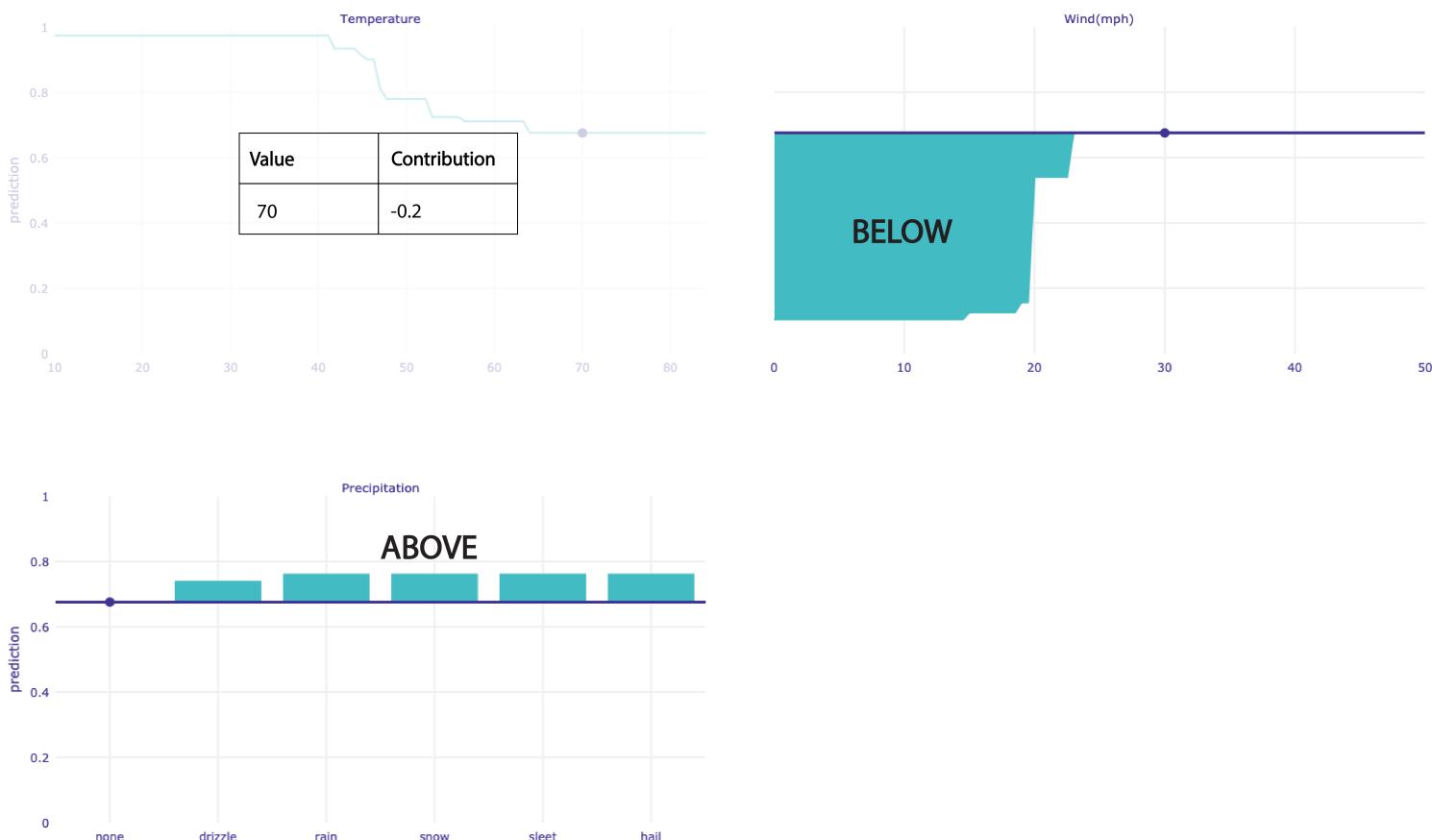
The **Contribution** value when you mouse over the Precipitation chart has a **negative (-)** sign. Also, all of the bars for the Precipitation chart are pointing up **above** the prediction dot. This means that changing precipitation is probably going to **increase** the probability that you should wear a coat. So the current value of precipitation is also pushing the model toward predicting 'NO'.

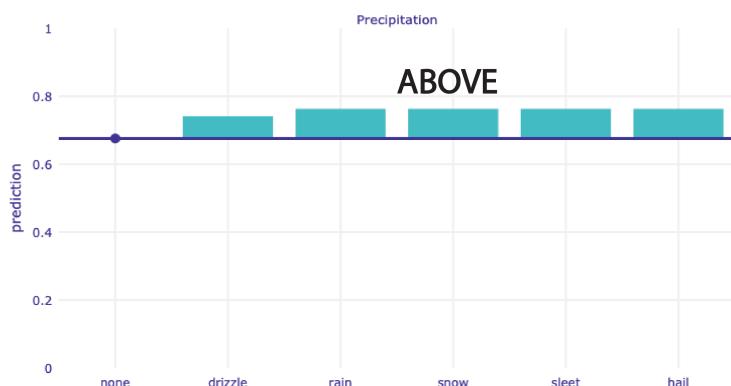
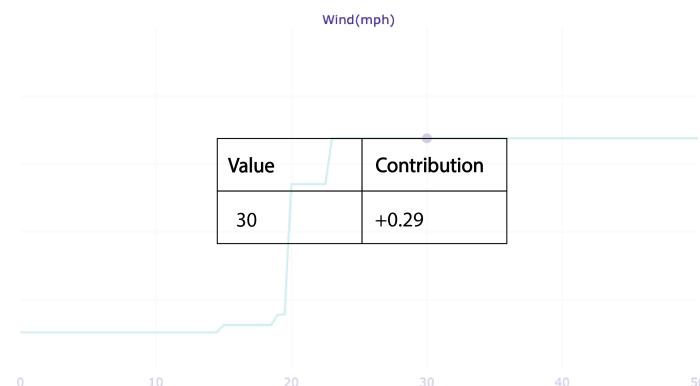
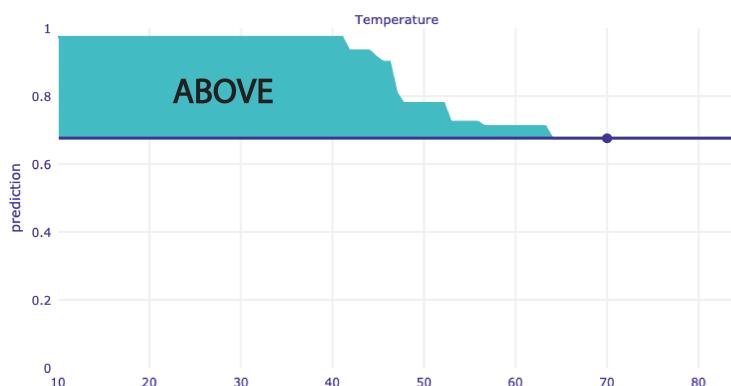
Meanwhile, the **Contribution** value when you mouse over the Wind(mph) chart has a **positive (+)** sign. The area **below** the current prediction line in the Wind(mph) chart is large, while the area above the prediction line is 0. Changing the value of wind(mph) is probably going to **decrease** the probability that you should wear a coat. So the current value of wind is pushing the model toward predicting 'YES'.

Model Prediction: 0.648

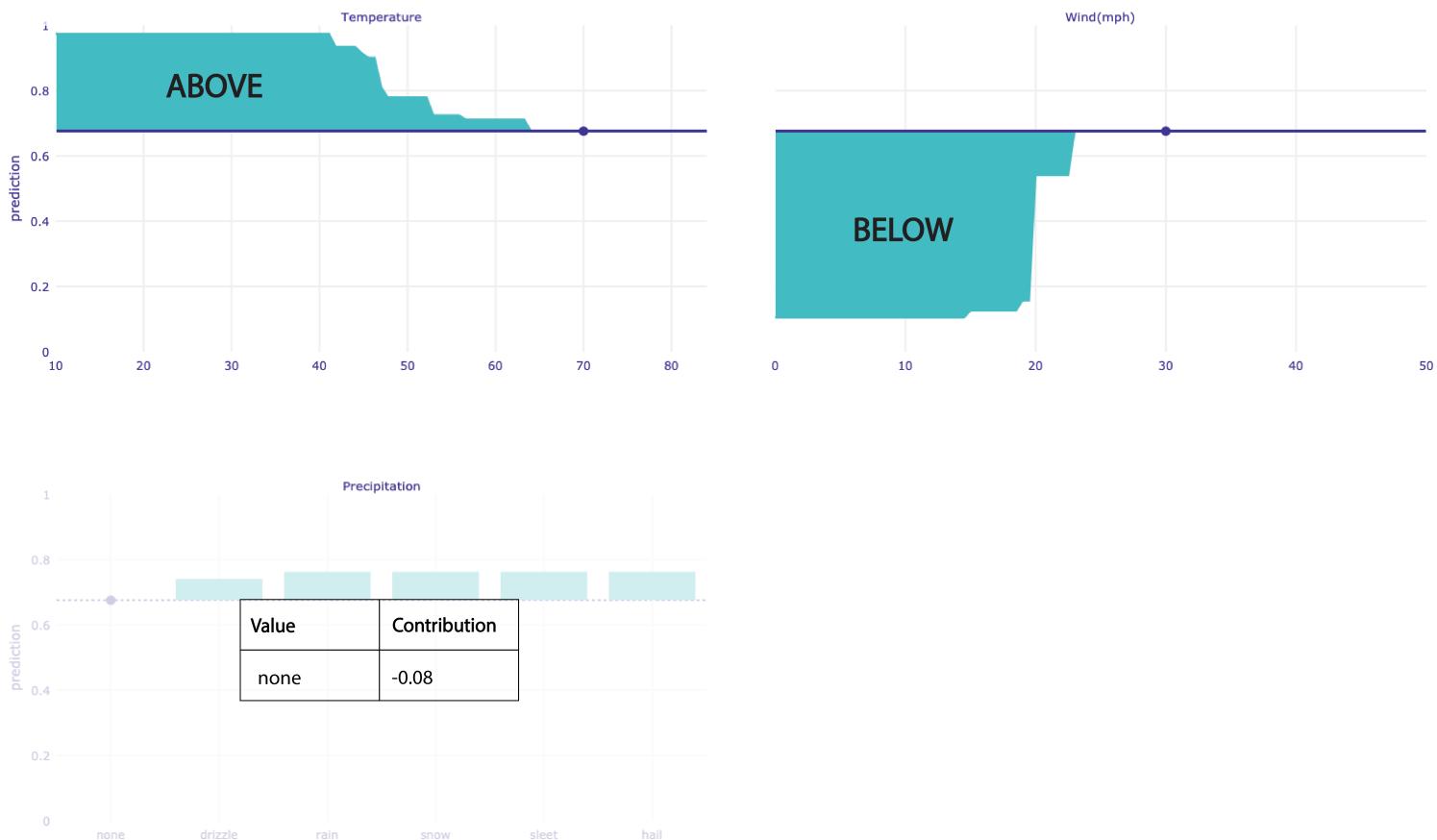


Model Prediction: 0.648



Model Prediction: 0.648

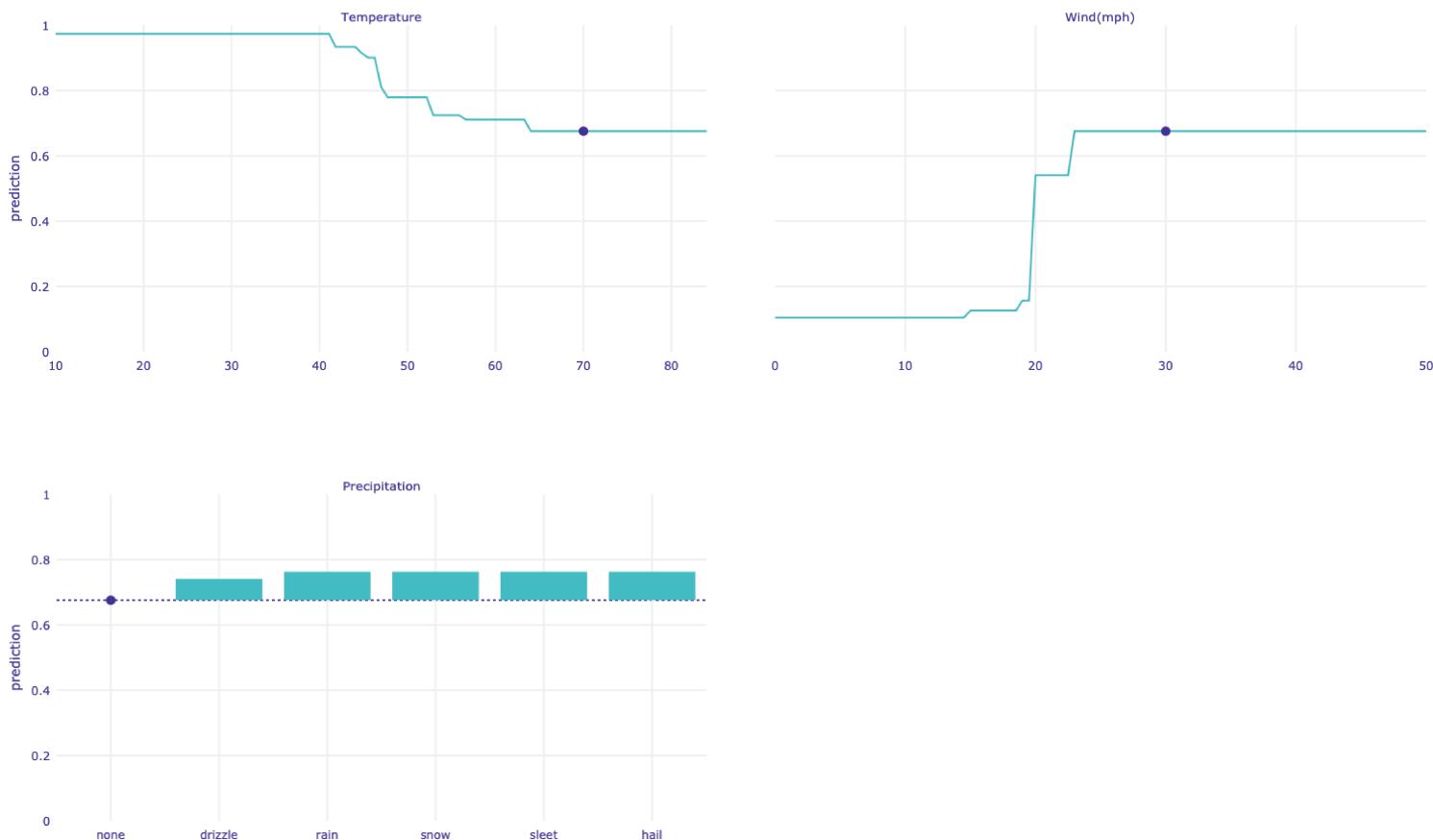
Model Prediction: 0.648



Which factor has the greatest predictive power?

- Temperature
- Wind
- Precipitation

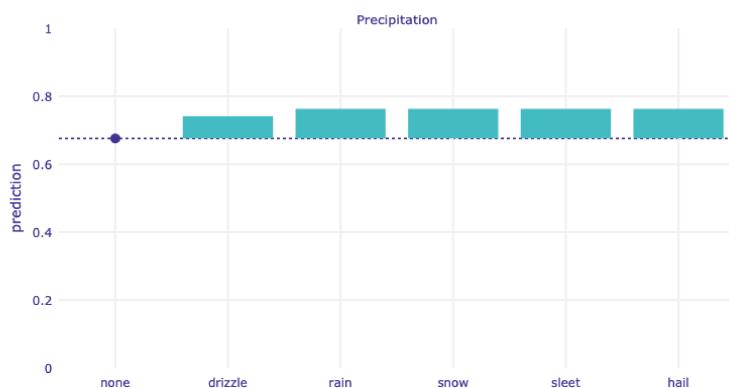
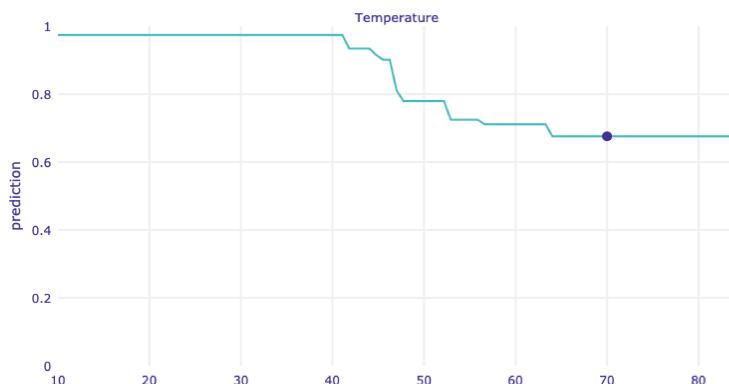
Model Prediction: 0.648



Model Prediction: 0.648



Model Prediction: 0.648



Model Prediction: 0.648



Correct –

The value in the **Contribution** column of the table you see when mousing over the Wind(mph) chart is 0.29, which is larger than 0.2 (Temperature) and 0.08 (Precipitation). This means Wind has the greatest predictive power.

Also, the area between the line and the current prediction in the Wind(mph) chart is larger than the area between the line and

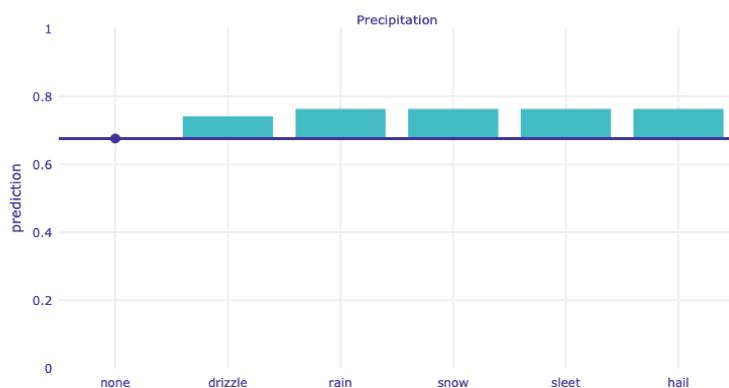
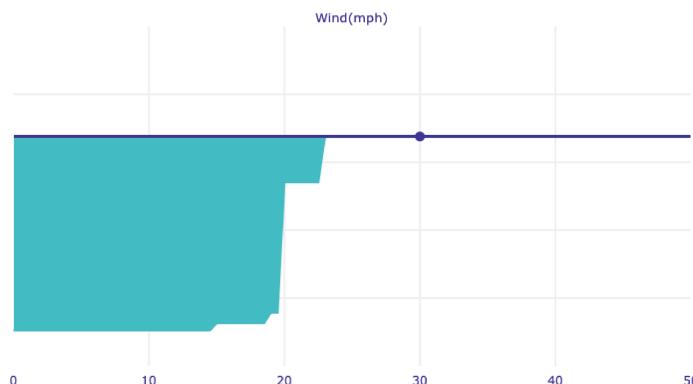
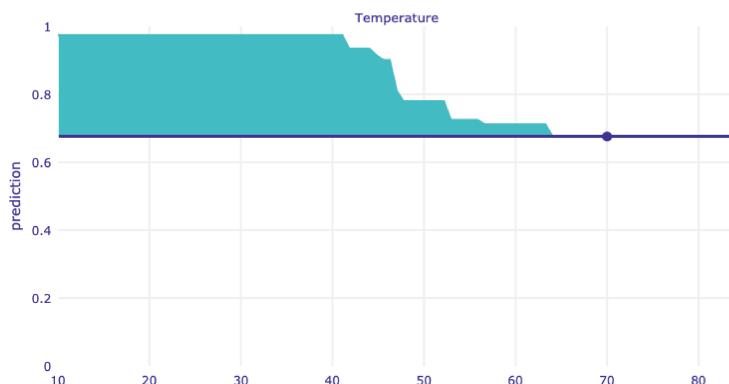
the current prediction in the Temperature chart, and the bars in the Precipitation chart. Changing the wind(mph) can cause a larger change in model output than changing the Temperature or the Precipitation.

Not quite -

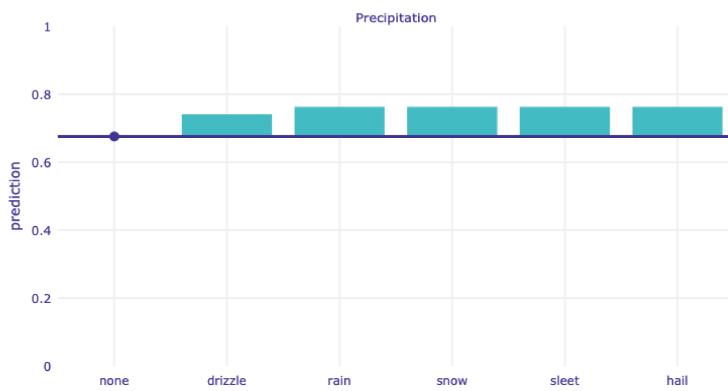
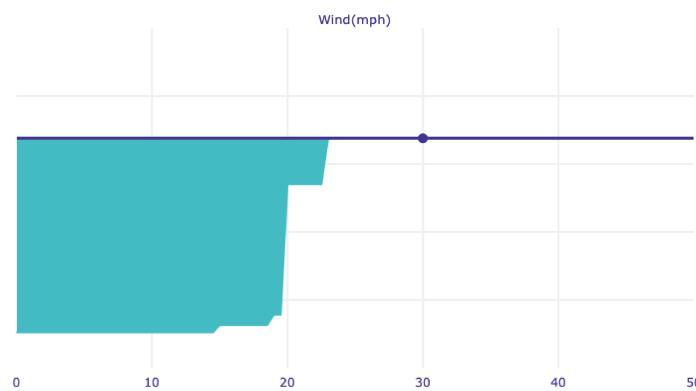
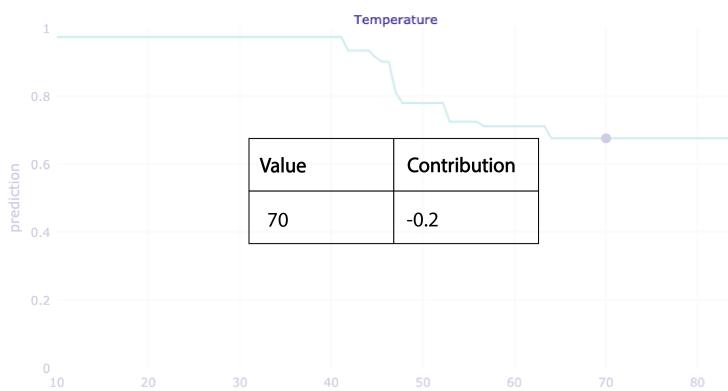
The value in the **Contribution** column of the table you see when mousing over the Wind(mph) chart is 0.29, which is larger than 0.2 (Temperature) and 0.08 (Precipitation). This means Wind has the greatest predictive power.

Also, the area between the line and the current prediction in the Wind(mph) chart is larger than the area between the line and the current prediction in the Temperature chart, and the bars in the Precipitation chart. Changing the wind(mph) can cause a larger change in model output than changing the Temperature or the Precipitation.

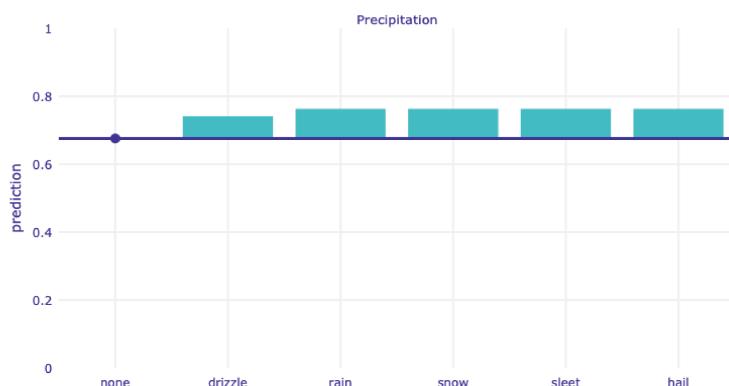
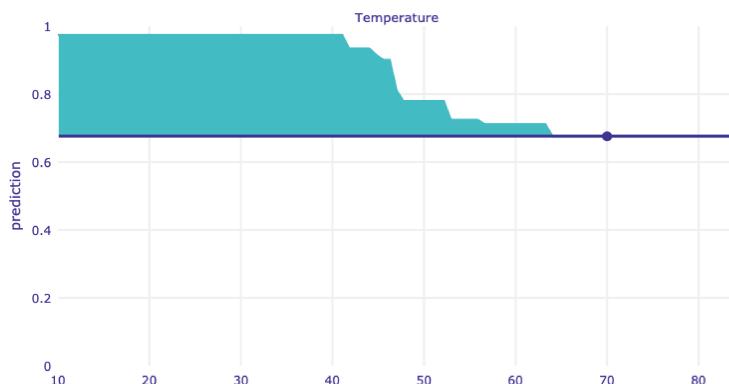
Model Prediction: 0.648

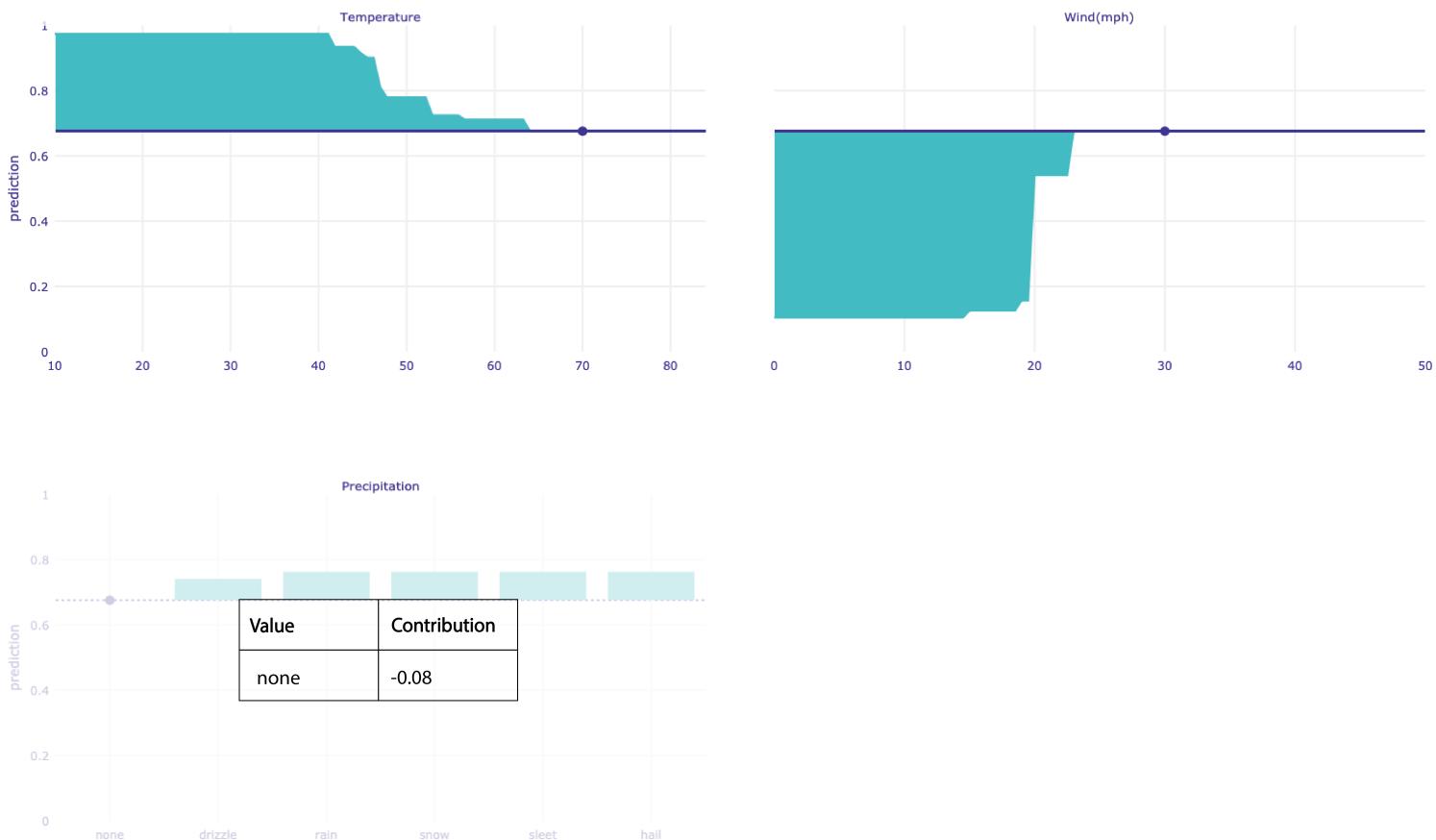


Model Prediction: 0.648



Model Prediction: 0.648



Model Prediction: 0.648

Intro Main

We have another machine learning model that makes predictions to approve or deny a loan based on a set of factors related to the loan applicant.

The model is trained to predict a person's likely income using real data from 26,000 people, and uses this prediction to decide

whether a person is likely to be able to pay back a loan. If the person is likely, the model outputs 'YES', they should be given a loan. If the person is not likely, the model outputs 'NO', they should not be given a loan.

The model generates a prediction based on each set of input values. If the predicted value is greater than or equal to 0.5, then the model will approve the loan. If the predicted value is less than 0.5, the model will deny the loan.

Six people applied to the loan. We input their corresponding values for each factor into the model.

We will show you six predictions the models generated for each of the six loan applicants.

Keep in mind that all six predictions were made by the **same** model.

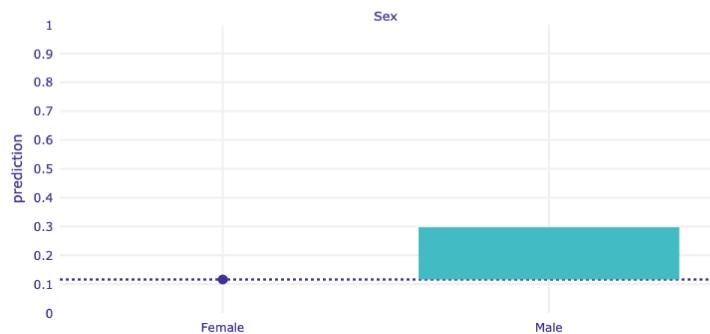
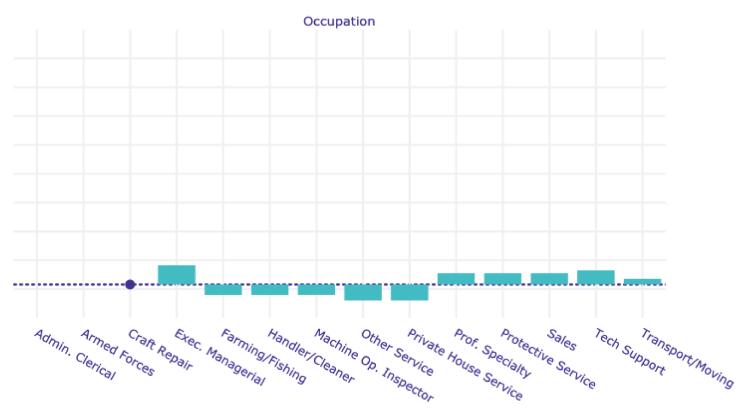
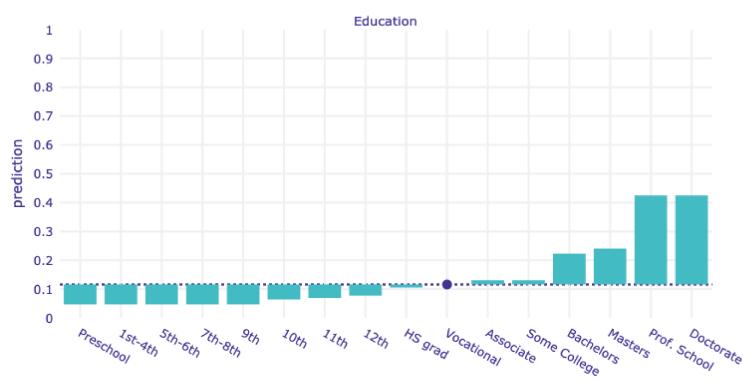
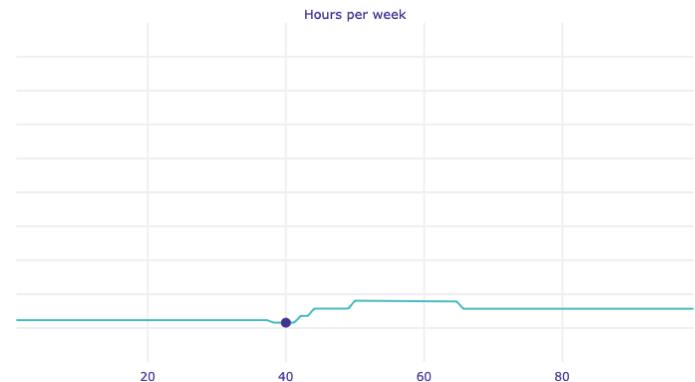
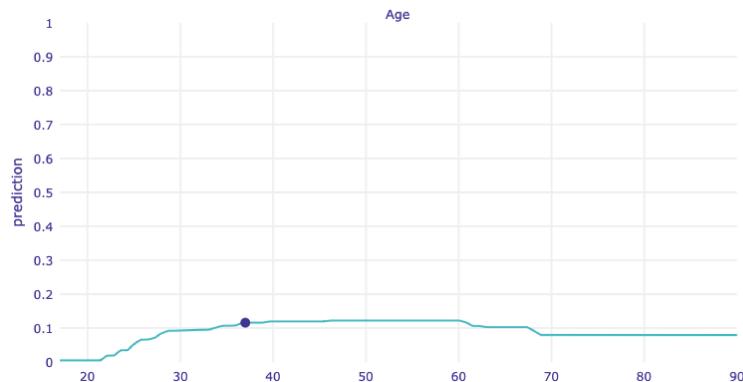
Woman 1

Below you will find the information of Applicant X.

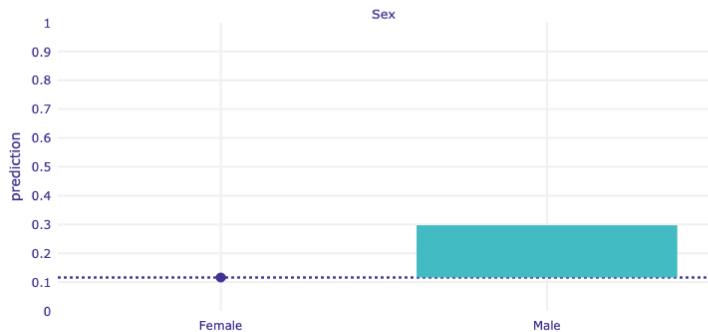
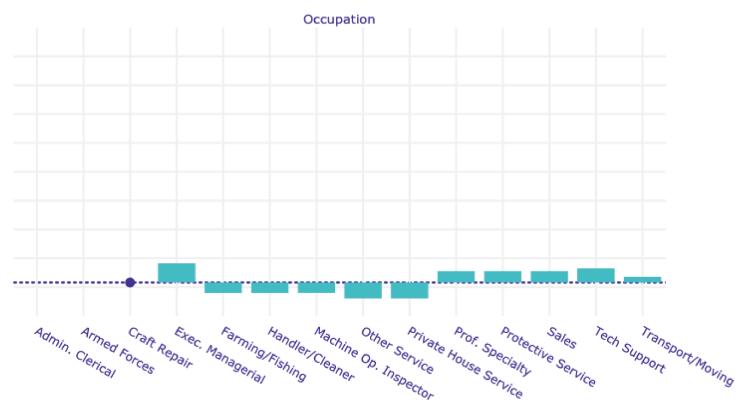
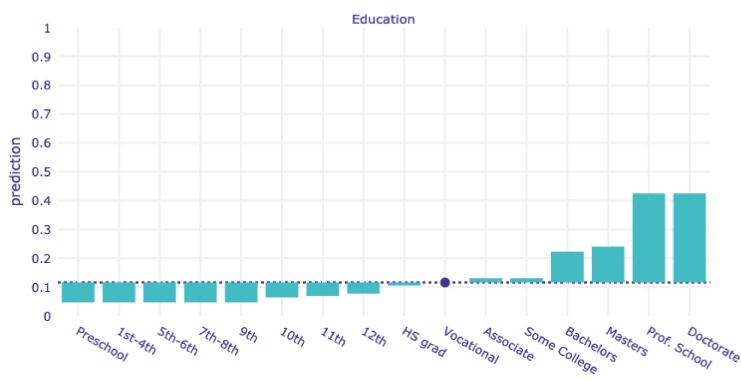
You can see that the model made a prediction of whether to approve or deny a loan from this applicant based on five factors. The explanation is below.

Look at the explanation, and answer the questions that follow.

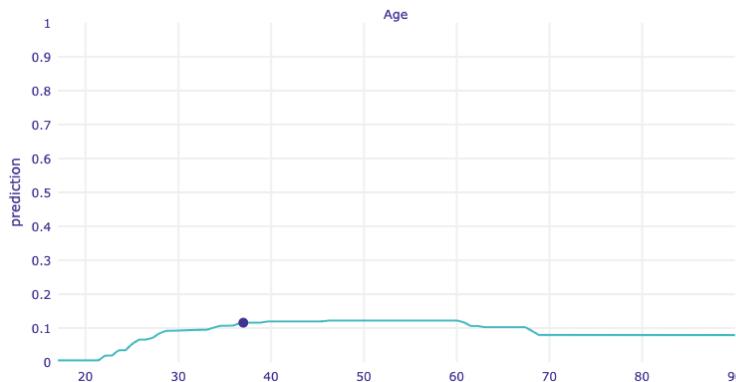
Remember that if the model's prediction probability (Predicted Value) for 'YES' is greater than or equal to 0.5, the model will return 'YES' (approve the loan). If it is less than 0.5, the model will return 'NO' (deny the loan).

Model Prediction: 0.116

Model Prediction: 0.116

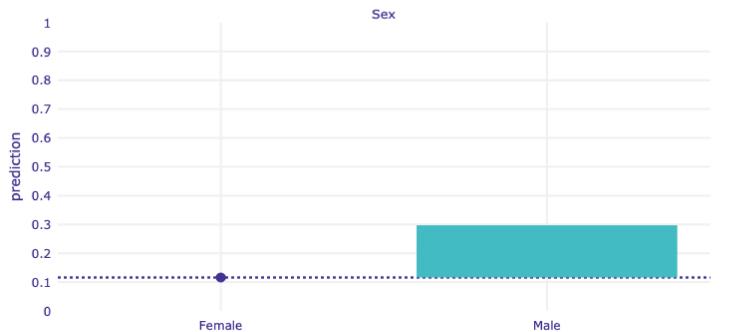
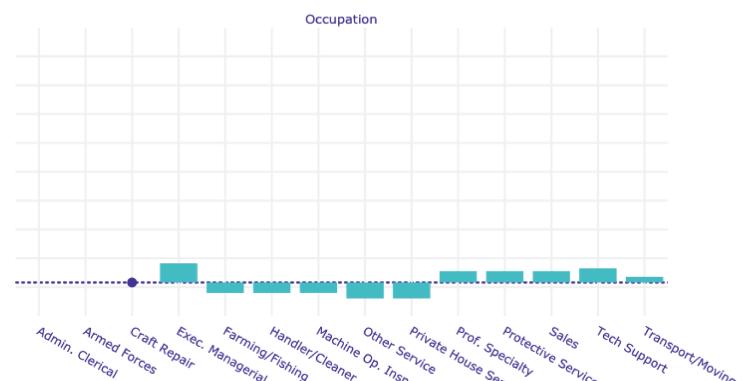


Model Prediction: 0.116

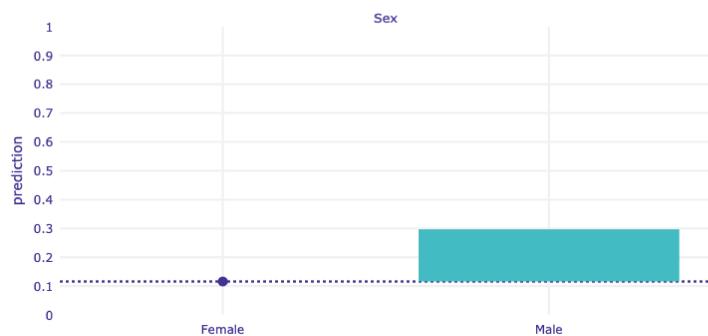
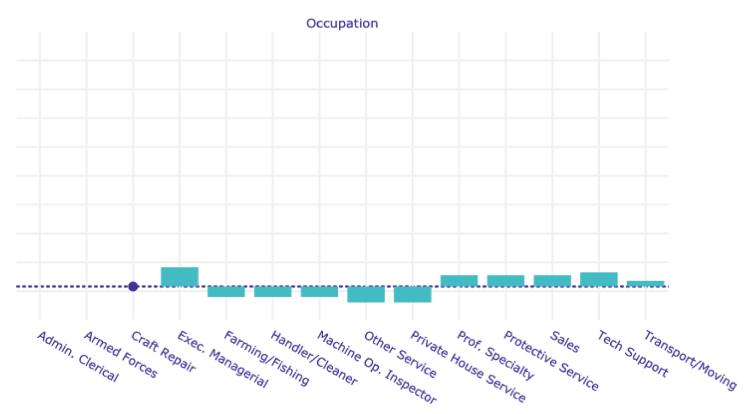
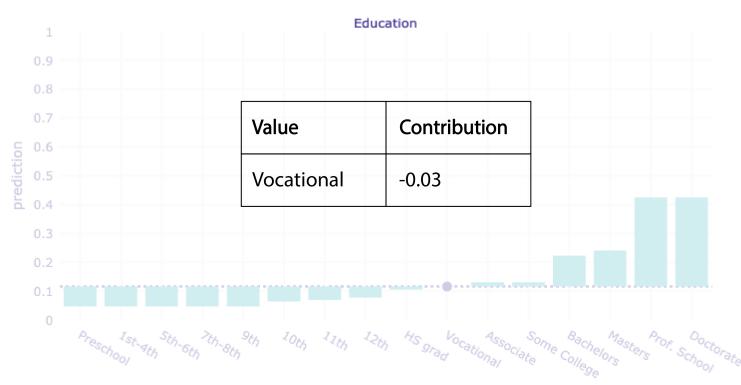
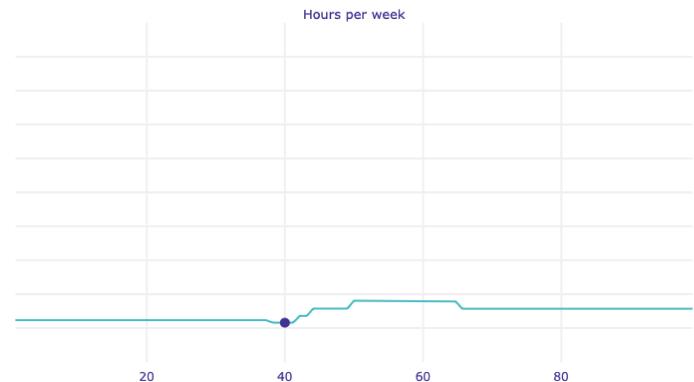
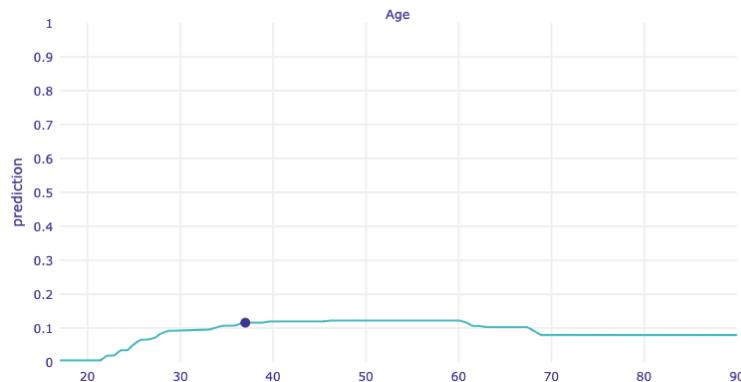


Hours per week

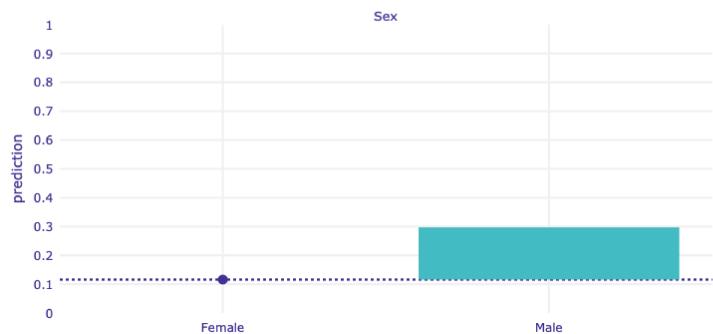
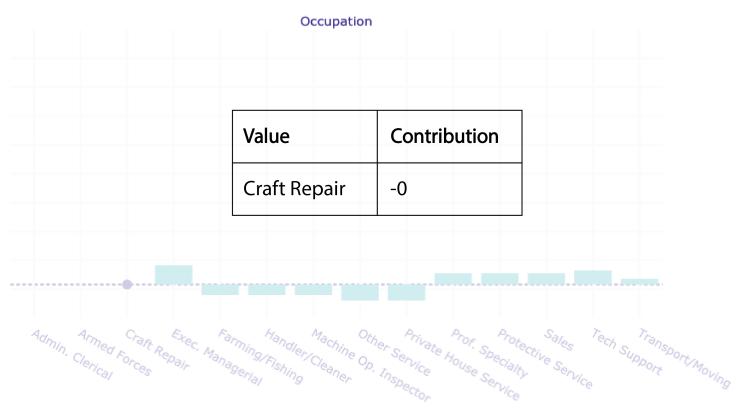
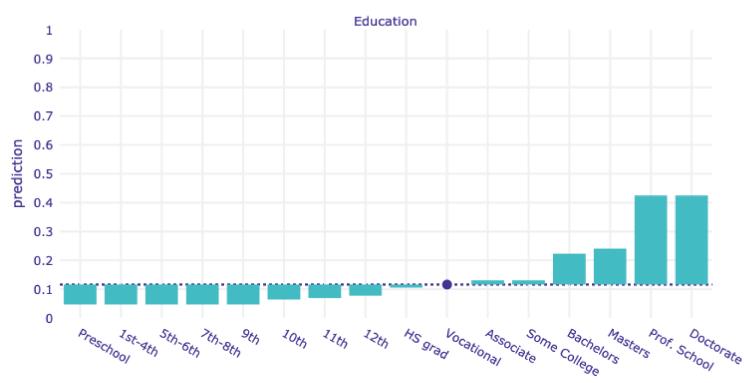
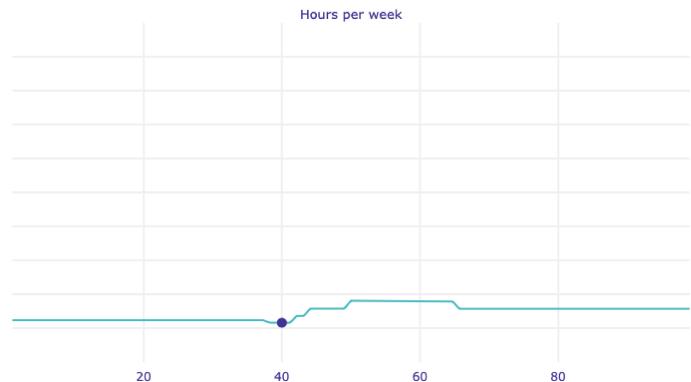
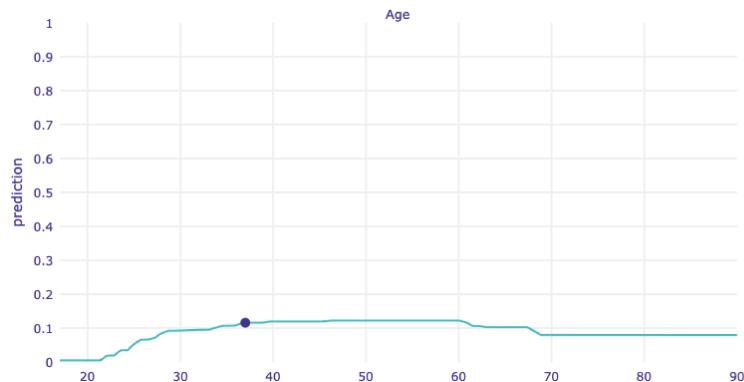
Value	Contribution
40	-0.04



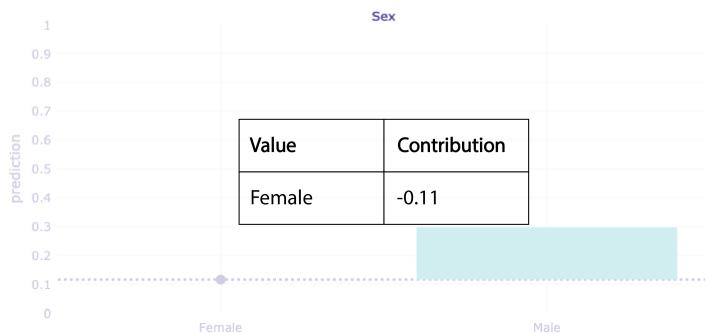
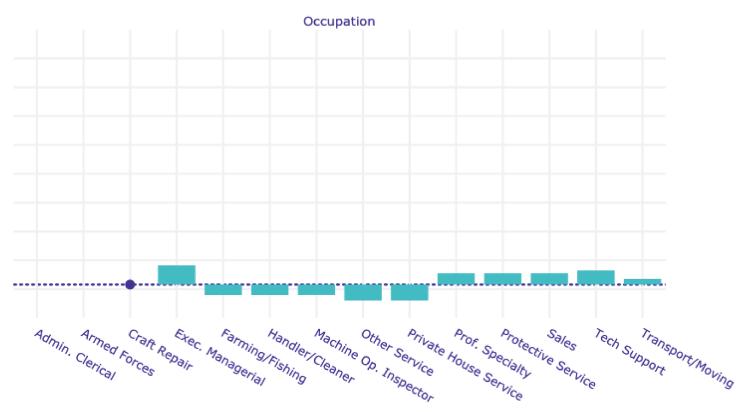
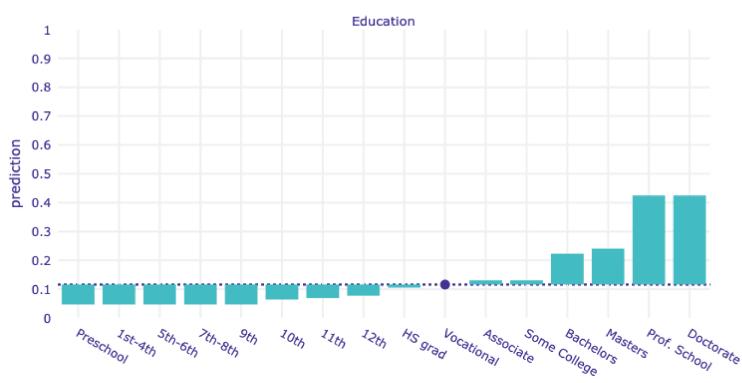
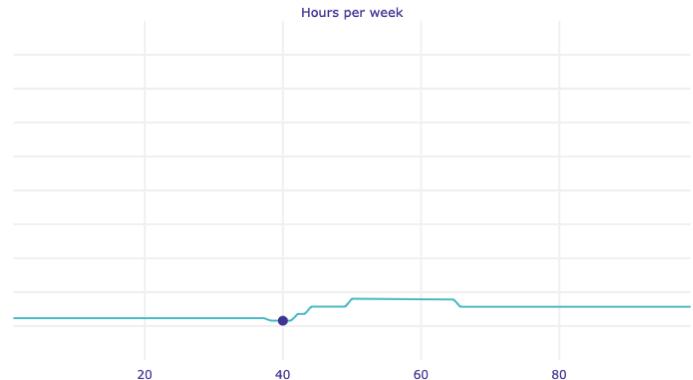
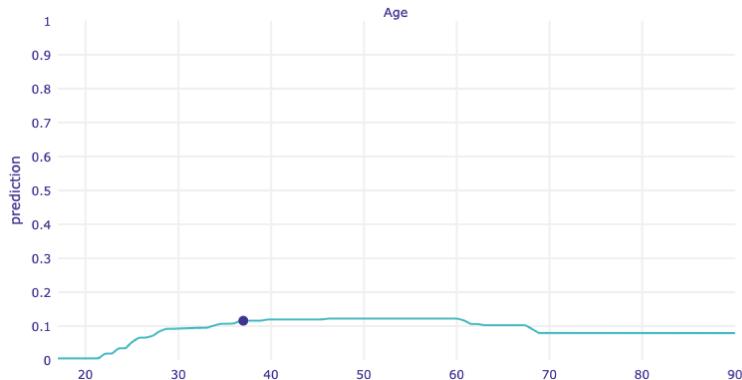
Model Prediction: 0.116



Model Prediction: 0.116



Model Prediction: 0.116



Will this model approve the loan for this person?

YES

NO

What feature was had the most predictive power for this decision?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation

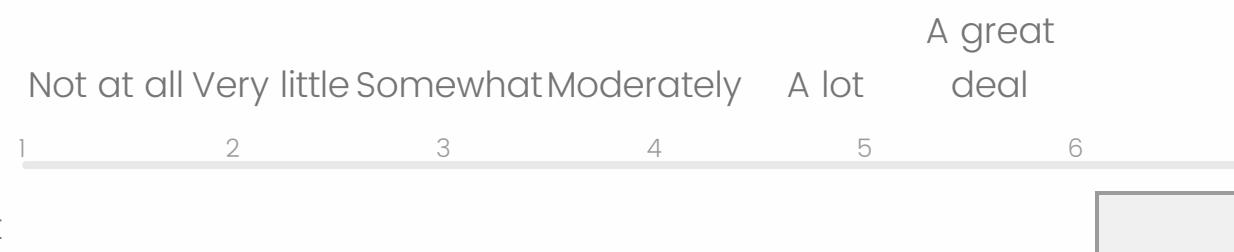
Which factor(s) are pushing the model toward predicting 'NO'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

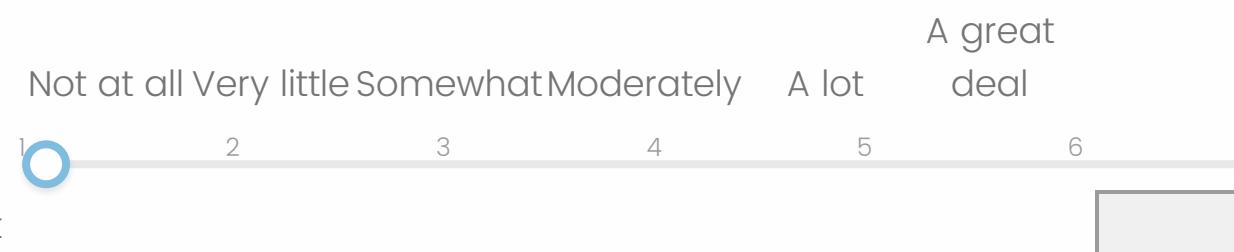
Which factor(s) are pushing the model toward predicting 'YES'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **you**?



On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **other people in general**?



Please indicate whether you agree with the below statements.

Agree

This model uses all of the features that it should use when making this decision.

This model does not use any unnecessary features when making this decision.

I trust the data this model was trained on.

Computer models can be trusted to make human decisions.

This model is accurate.

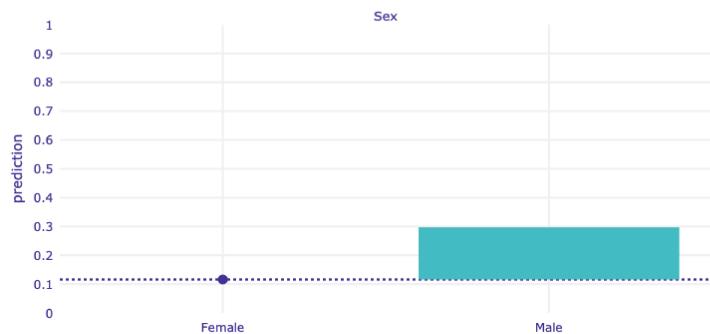
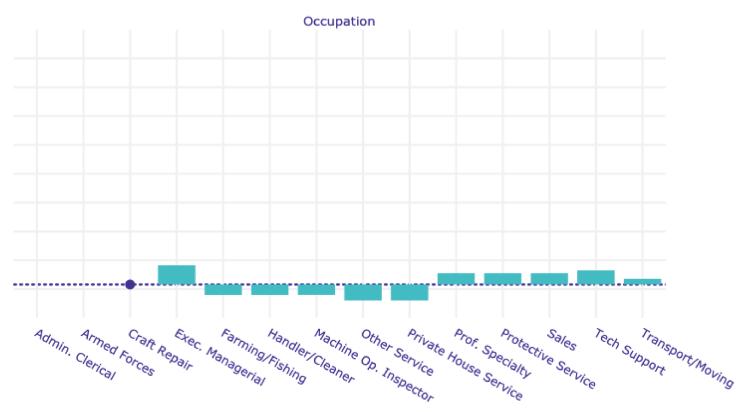
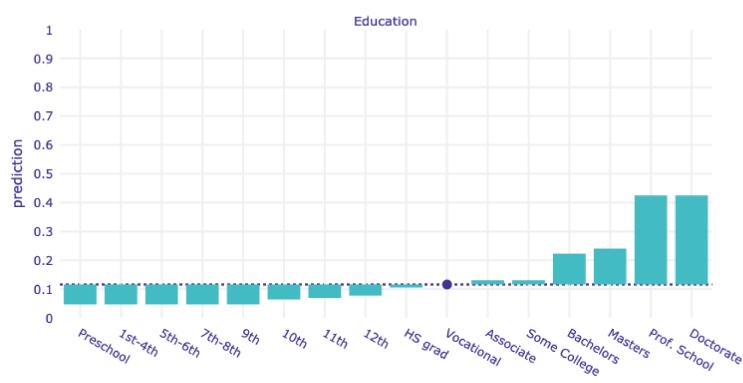
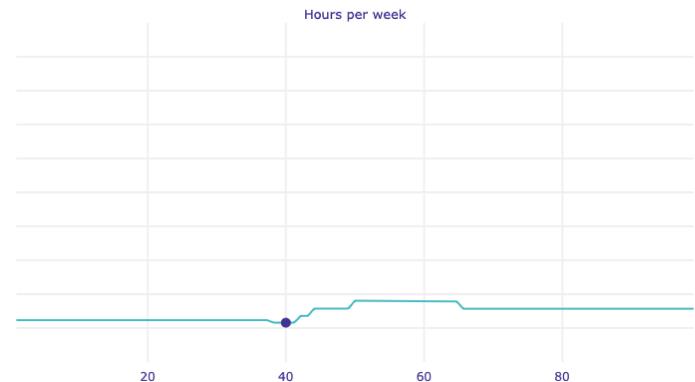
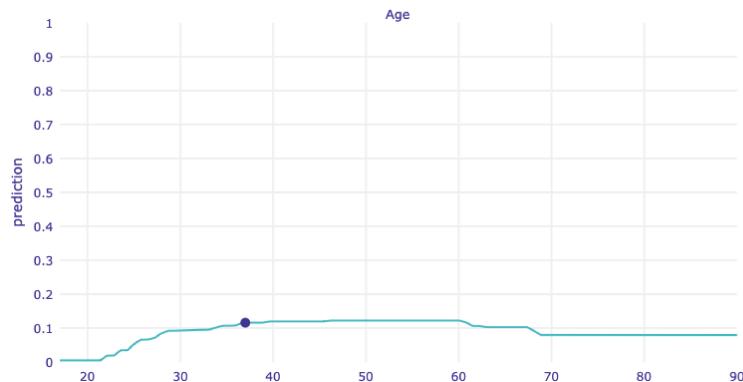
This model is fair.

This model would probably give me a loan because I am similar to the person described in this question.

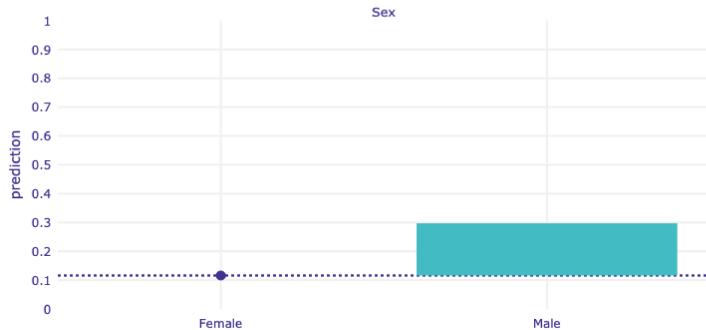
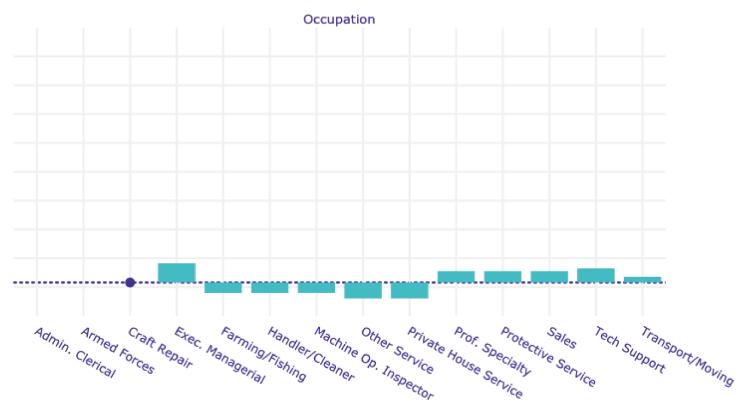
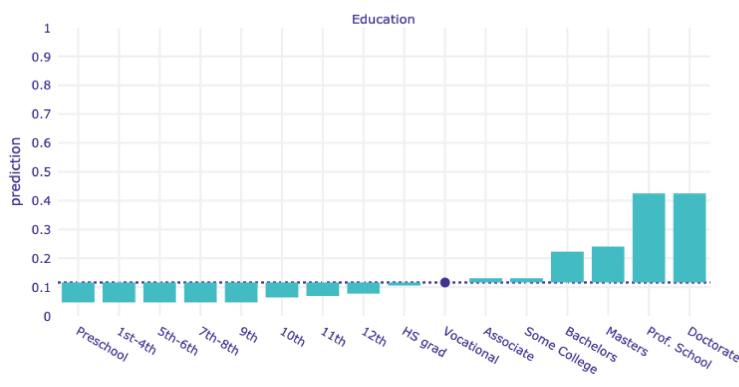
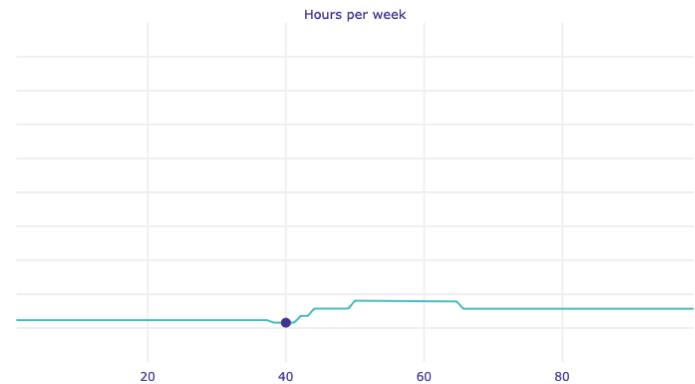
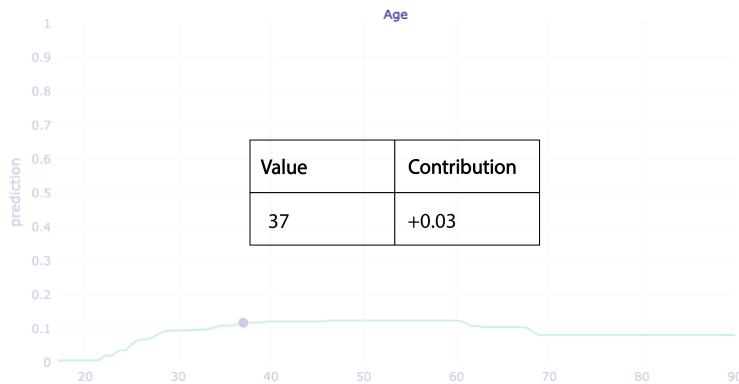
This model would probably give me a loan because I am different from the person described in this question.

This model would probably give me a loan because of previous decisions it has made.

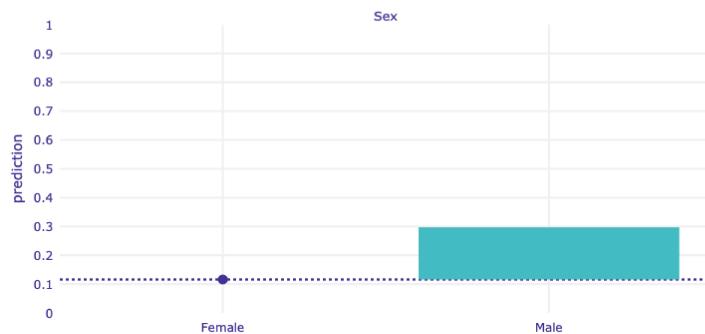
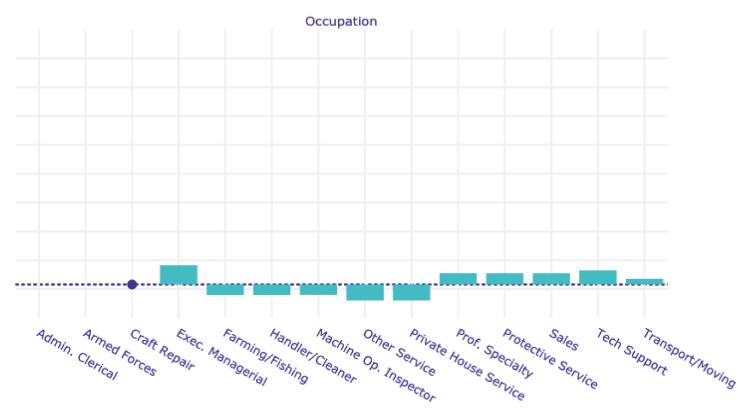
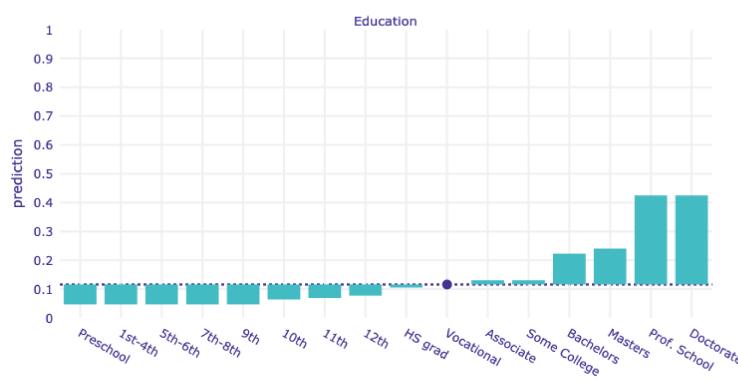
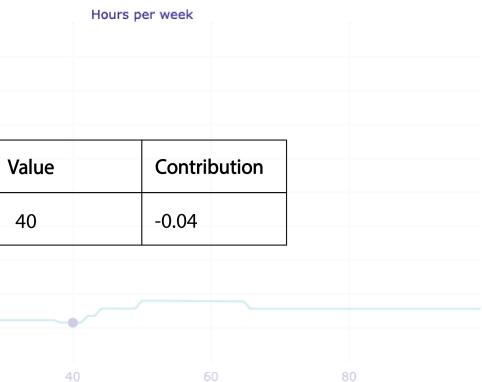
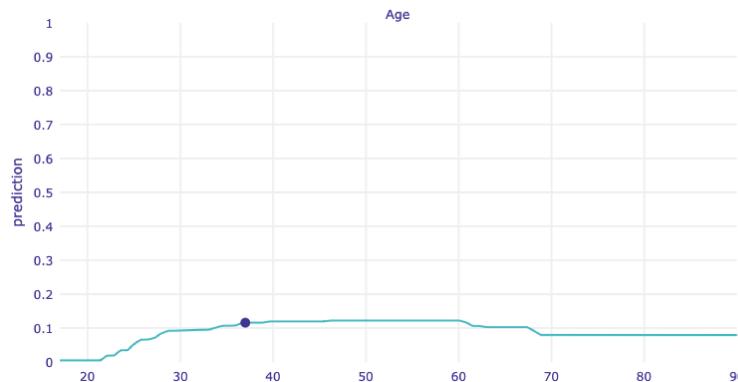
This model probably would not give me a loan, and this would be the correct decision.

Model Prediction: 0.116

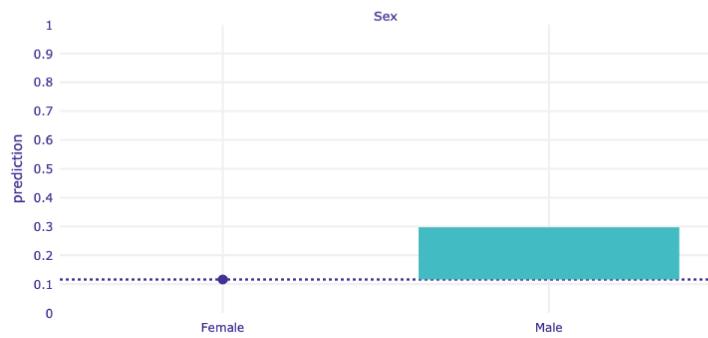
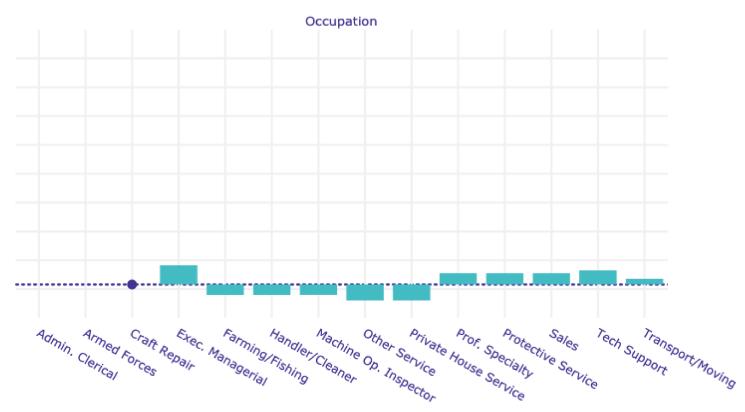
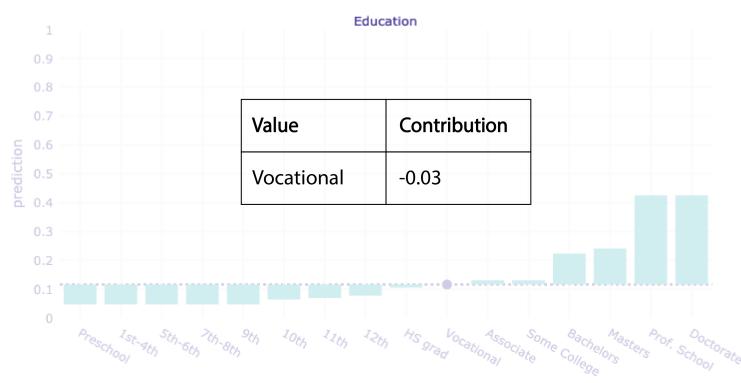
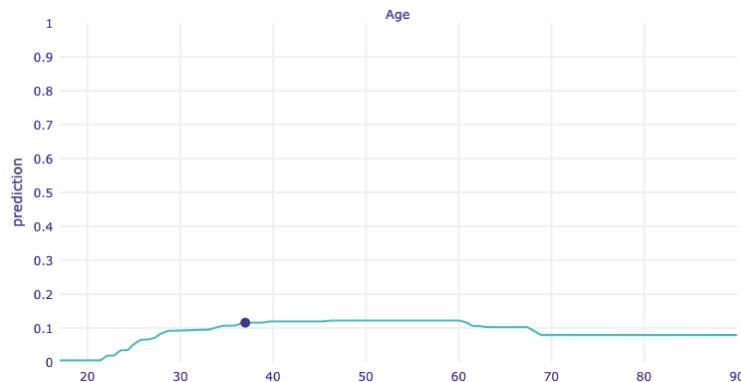
Model Prediction: 0.116



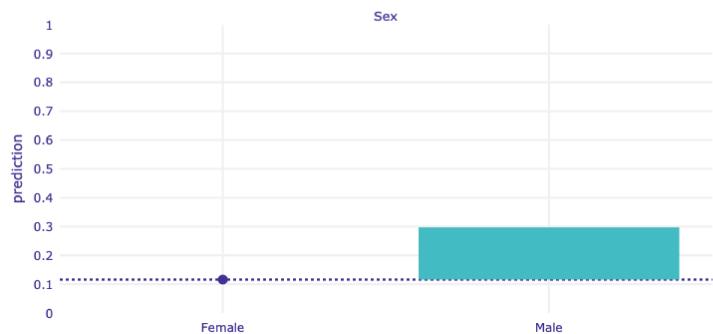
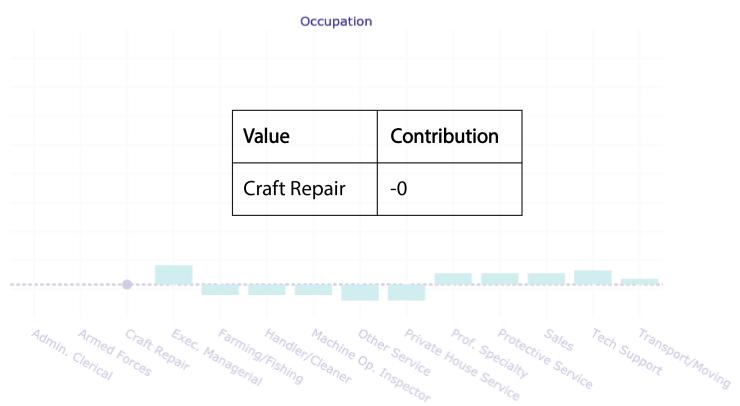
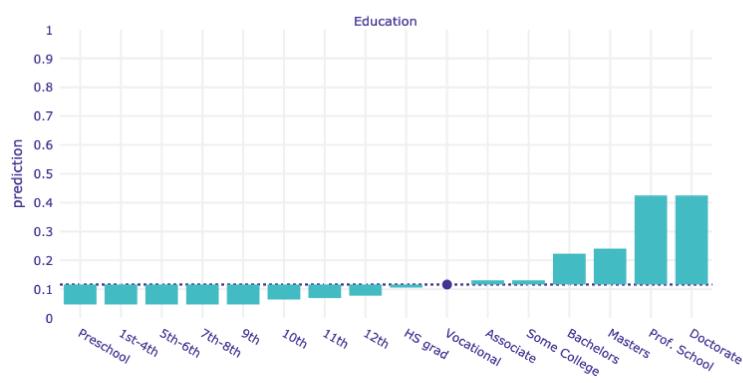
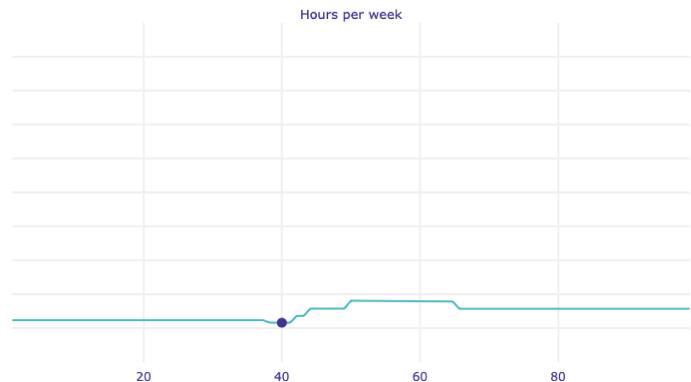
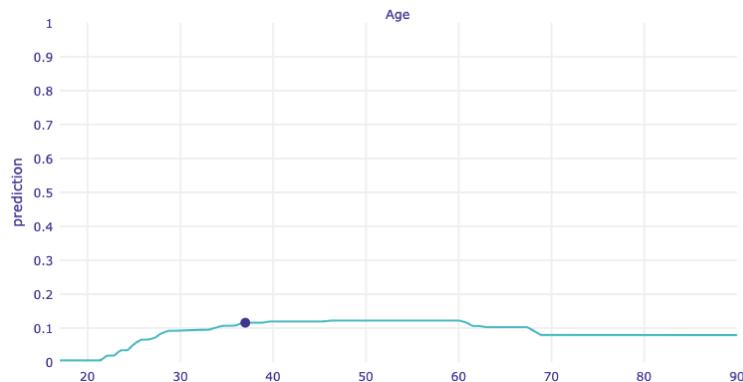
Model Prediction: 0.116



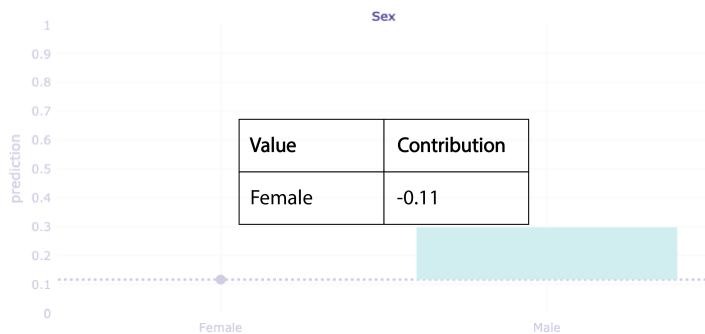
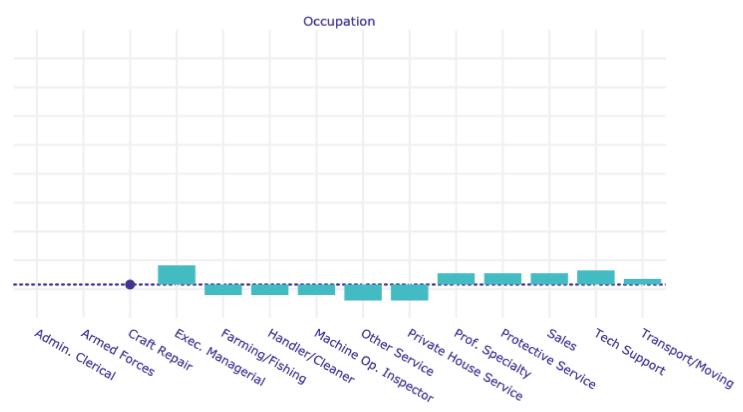
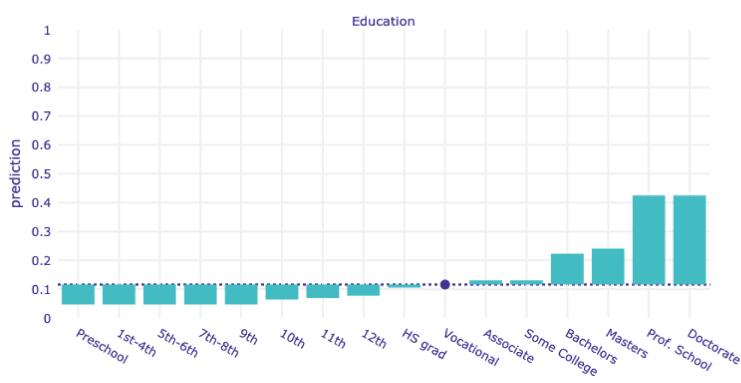
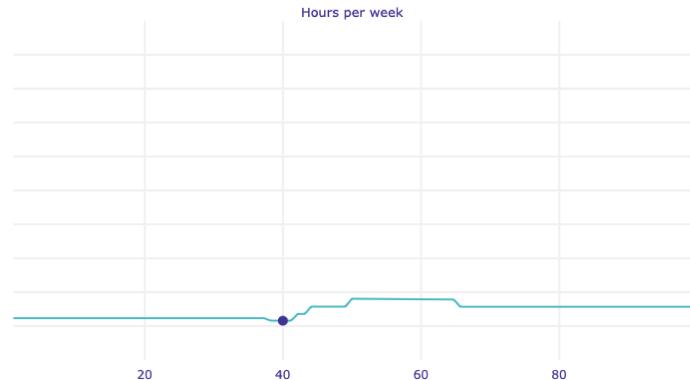
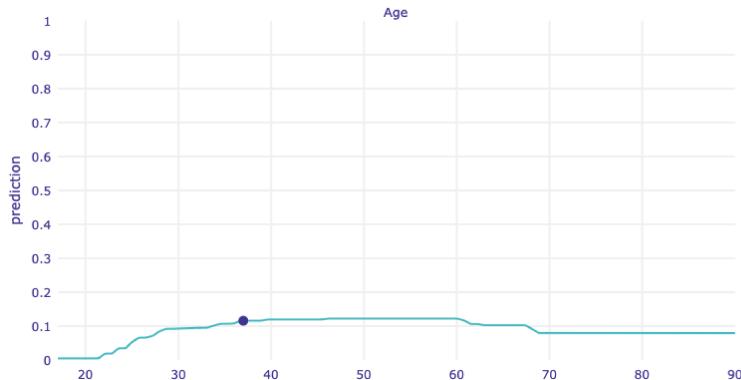
Model Prediction: 0.116



Model Prediction: 0.116



Model Prediction: 0.116



When answering the previous questions about the given explanation, which design aspects of the visualization did you find **most** useful?

When answering the previous questions about the given explanation, which design aspects of the visualizations did you find **least** useful?

Woman 2

Below you will find the information of Applicant R.

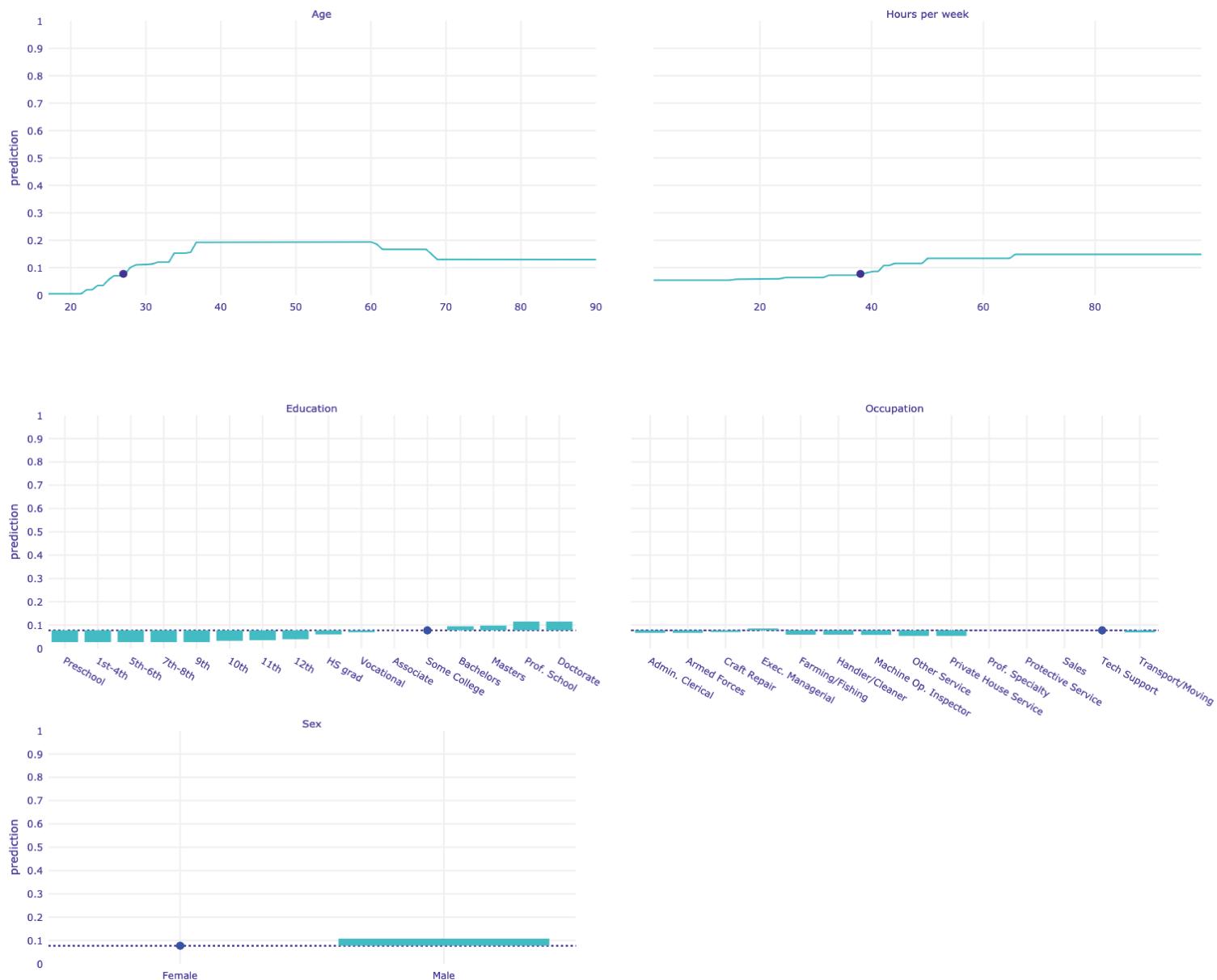
You can see that the model made a prediction of whether to approve or deny a loan from this applicant based on five factors. The explanation is below.

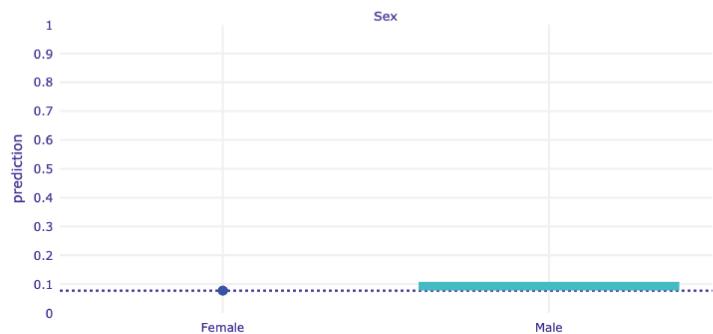
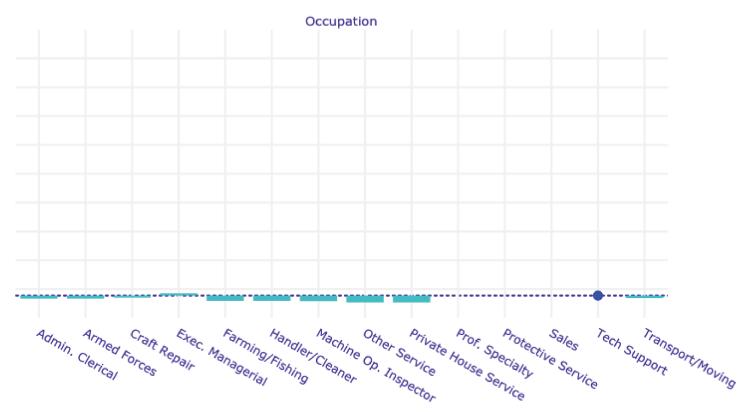
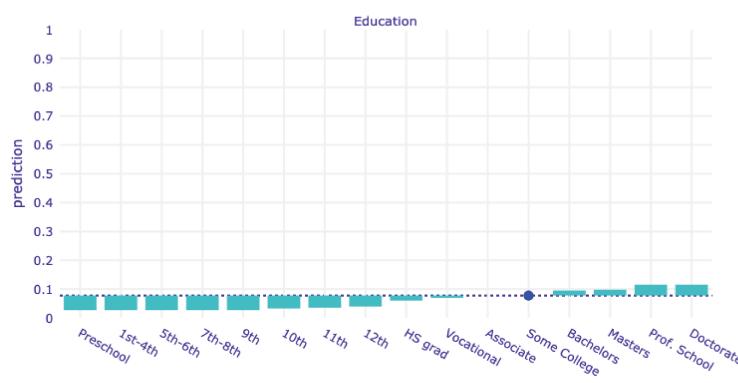
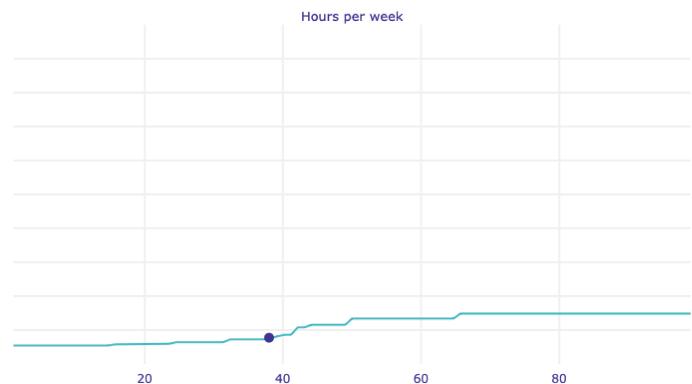
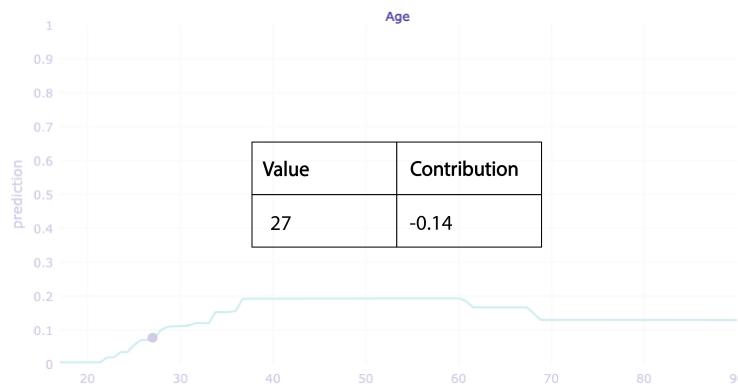
Look at the explanation, and answer the questions that follow.

Remember that if the model's prediction probability (Predicted Value) for 'YES' is greater than or equal to 0.5, the model will return 'YES' (approve the loan). If it is less than 0.5, the model will

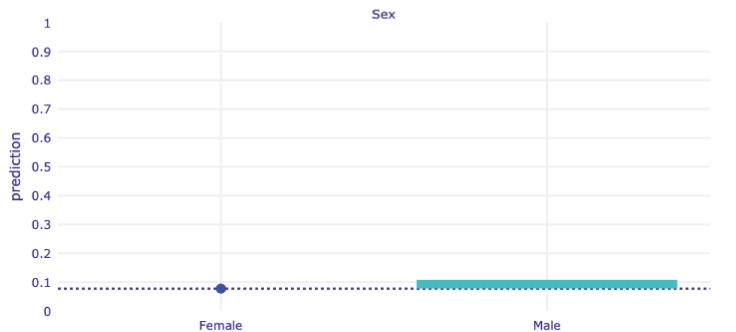
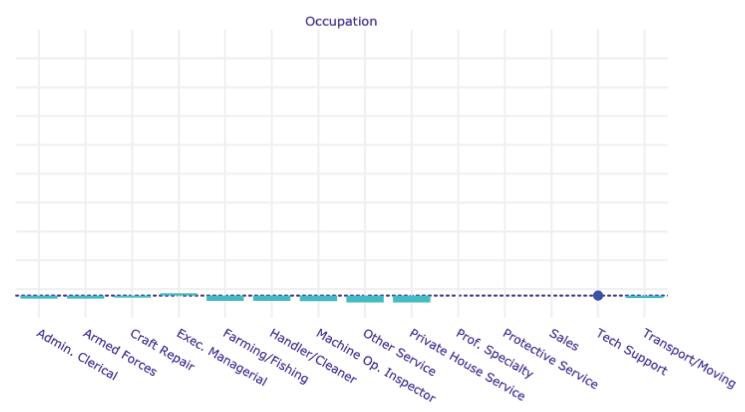
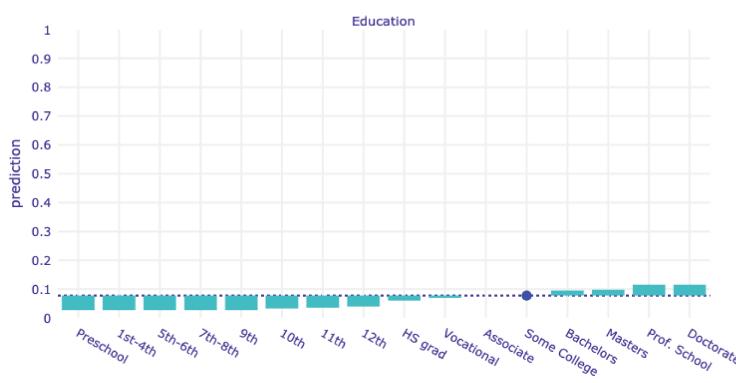
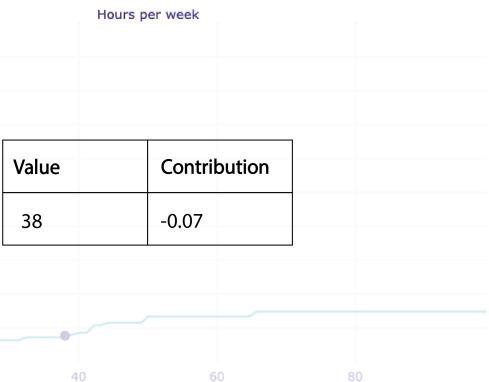
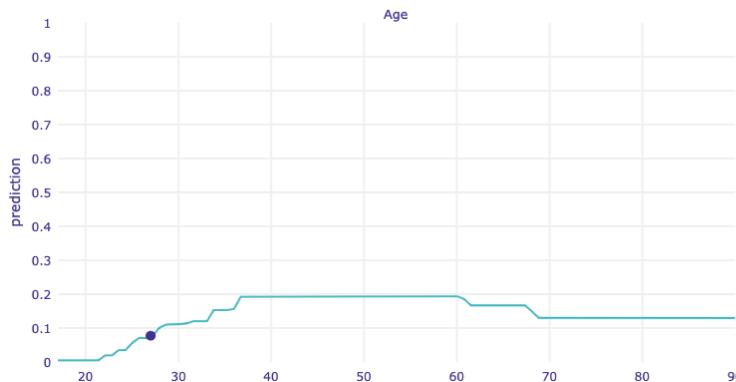
return 'NO' (deny the loan).

Model Prediction: 0.078

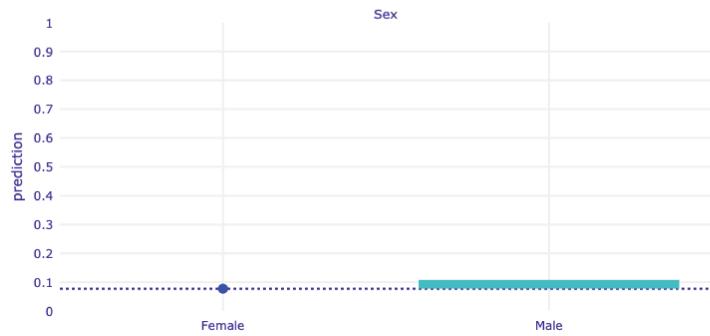
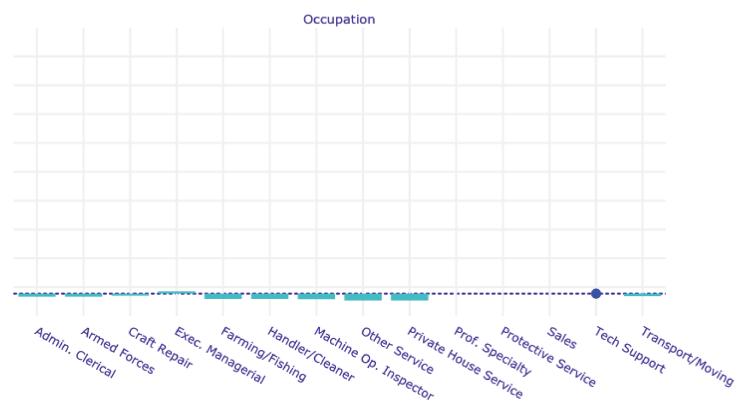
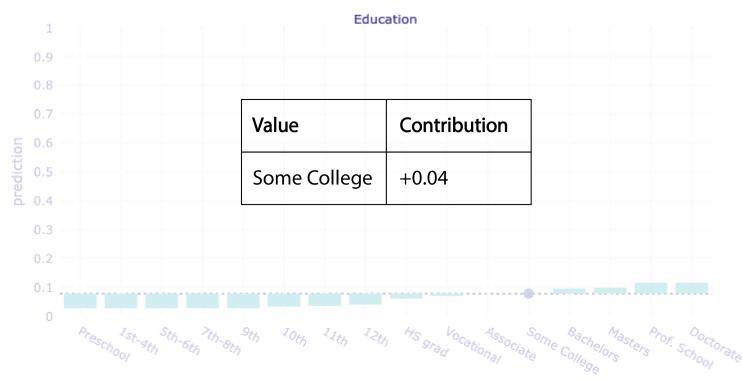
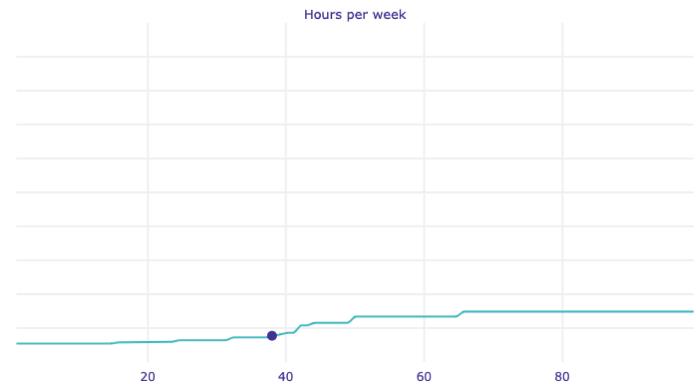
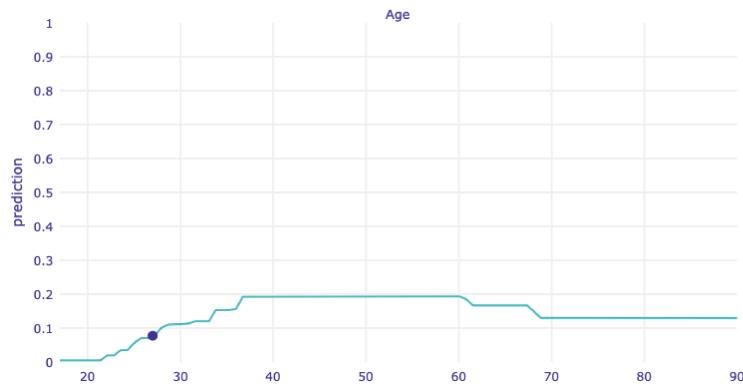


Model Prediction: 0.078

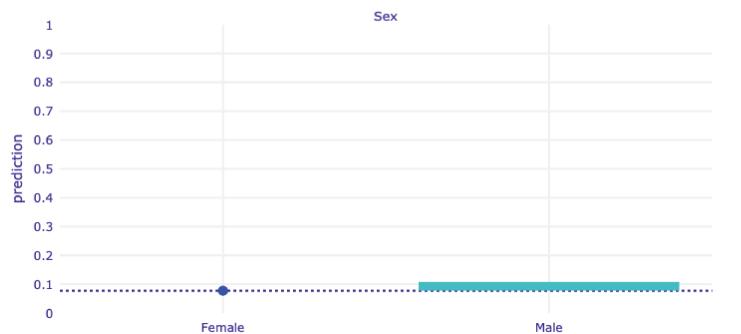
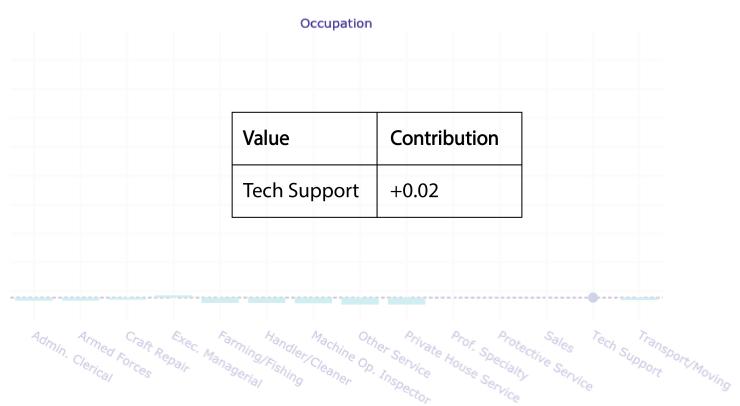
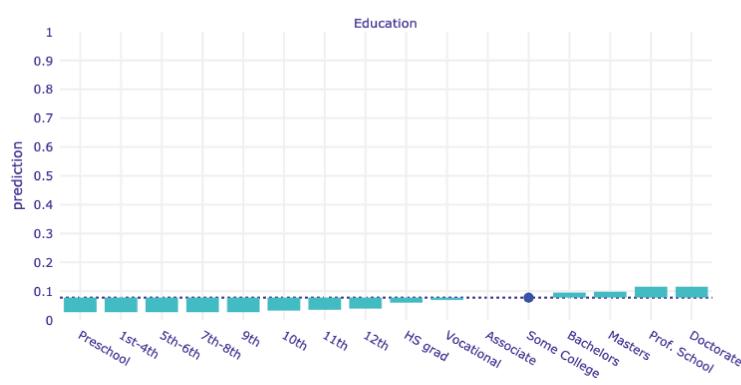
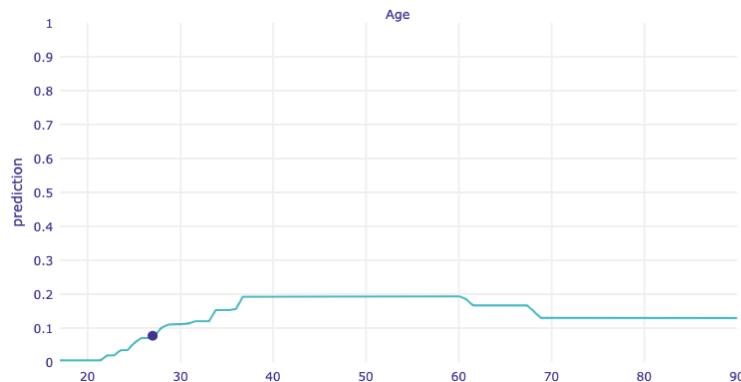
Model Prediction: 0.078



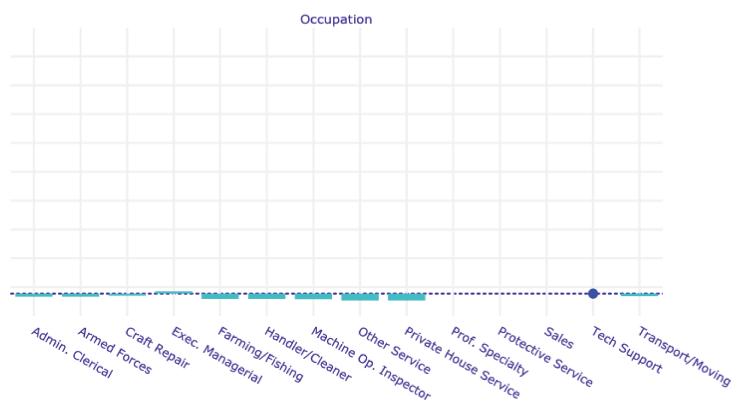
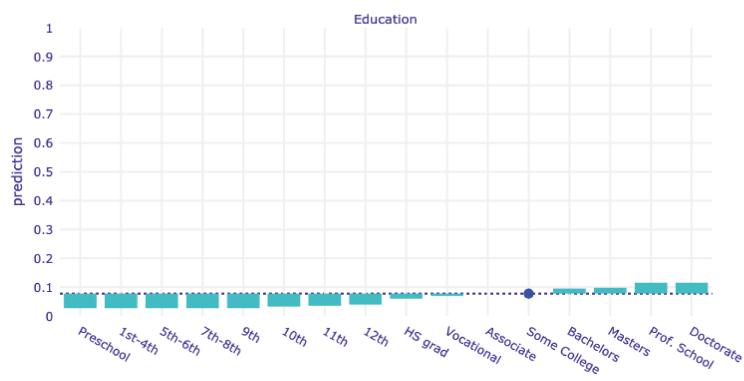
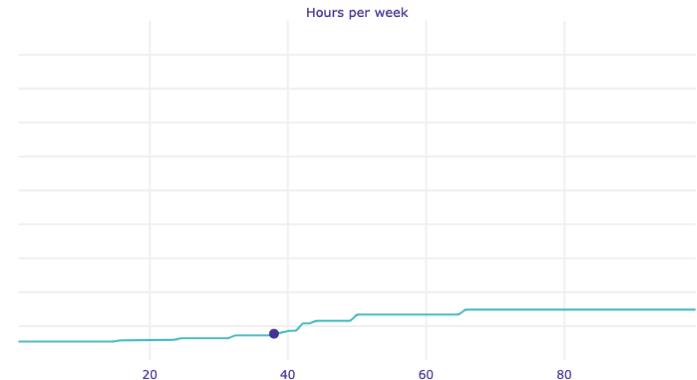
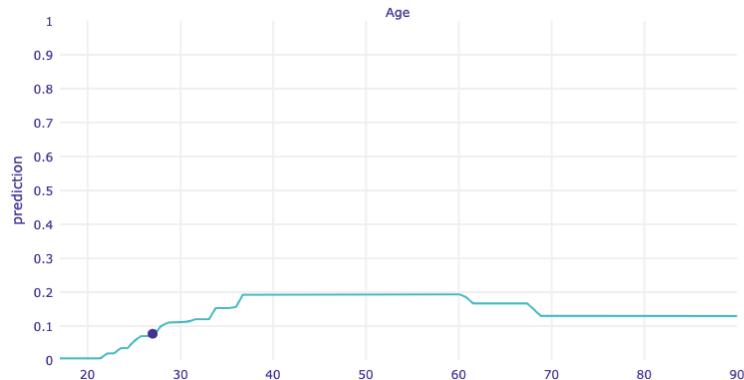
Model Prediction: 0.078



Model Prediction: 0.078



Model Prediction: 0.078



Will this model approve the loan for this person?

YES

NO

Which feature was had the most predictive power for this decision?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation

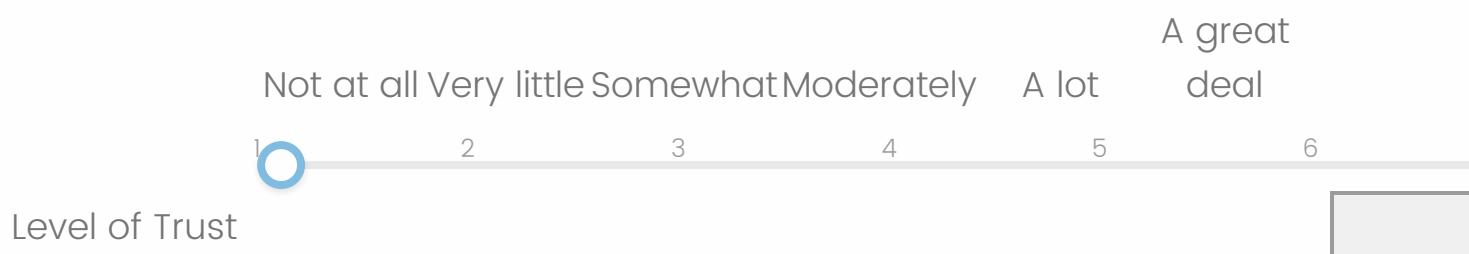
Which factor(s) are pushing the model toward predicting 'NO'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

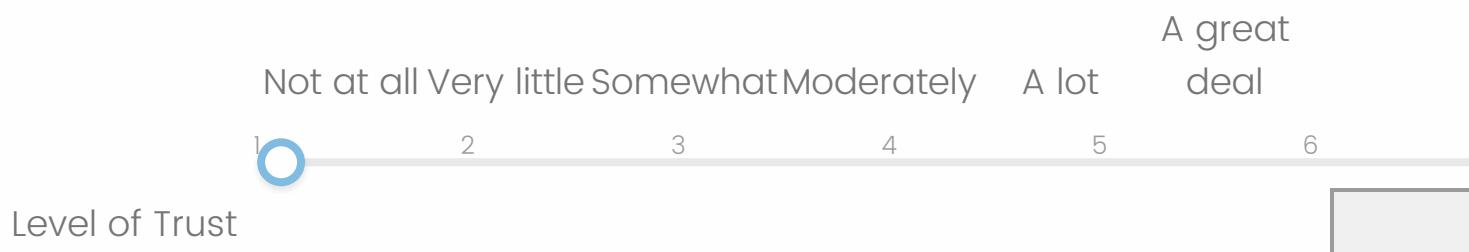
Which factor(s) are pushing the model toward predicting 'YES'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **you**?



On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **other people in general**?



Please indicate whether you agree with the below statements.

Agree

This model uses all of the features that it should use when making this decision.

This model does not use any unnecessary features when making this decision.

I trust the data this model was trained on.

Computer models can be trusted to make human decisions.

This model is accurate.

This model is fair.

This model would probably give me a loan because I am similar to the person described in this question.

This model would probably give me a loan because I am different from the person described in this question.

This model would probably give me a loan because of previous decisions it has made.

This model probably would not give me a loan, and this would be the correct decision.

Woman 3

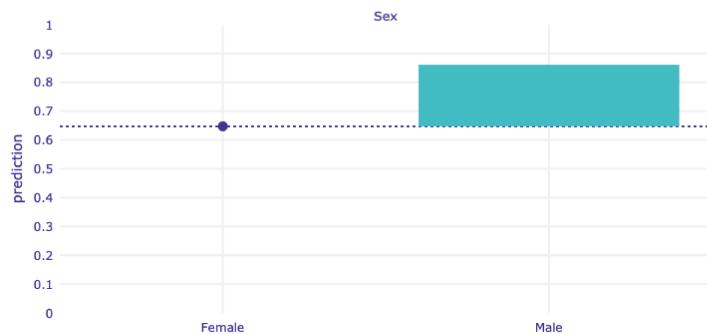
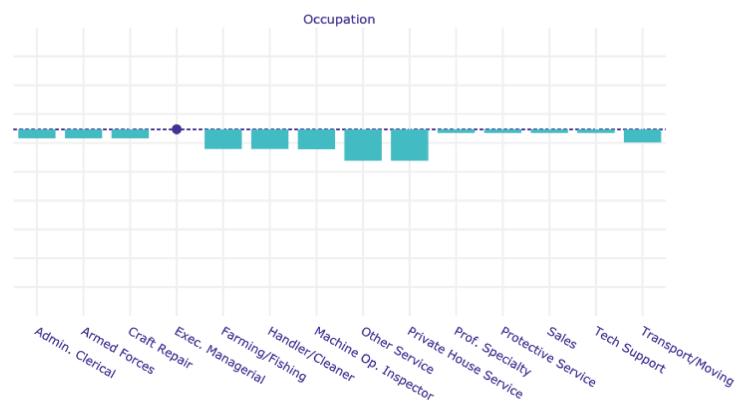
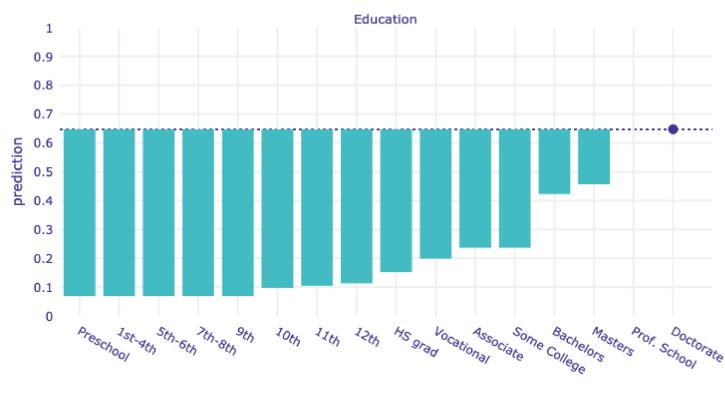
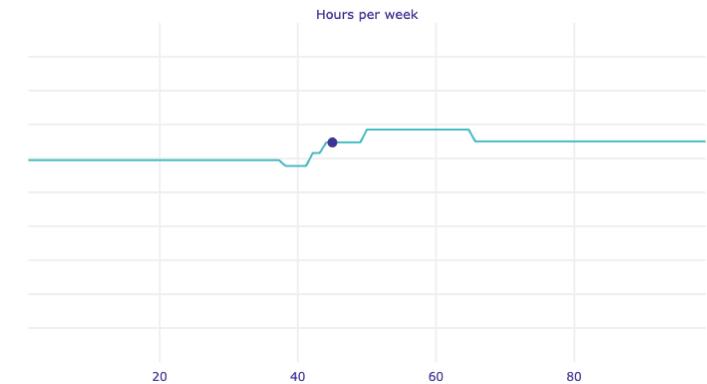
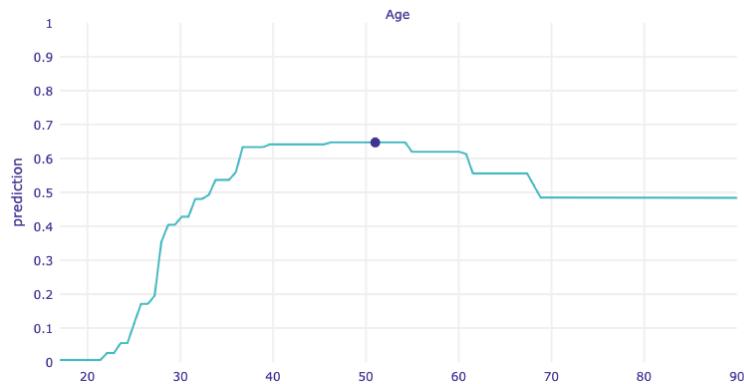
Below you will find the information of Applicant S.

You can see that the model made a prediction of whether to approve or deny a loan from this applicant based on five factors. The explanation is below.

Look at the explanation, and answer the questions that follow.

Remember that if the model's prediction probability (Predicted Value) for 'YES' is greater than or equal to 0.5, the model will return 'YES' (approve the loan). If it is less than 0.5, the model will return 'NO' (deny the loan).

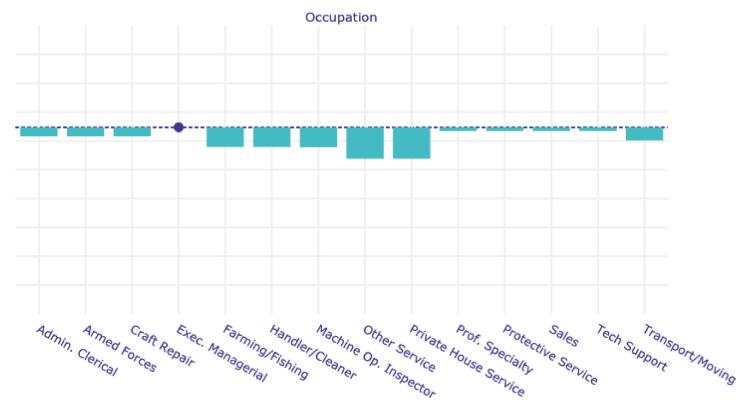
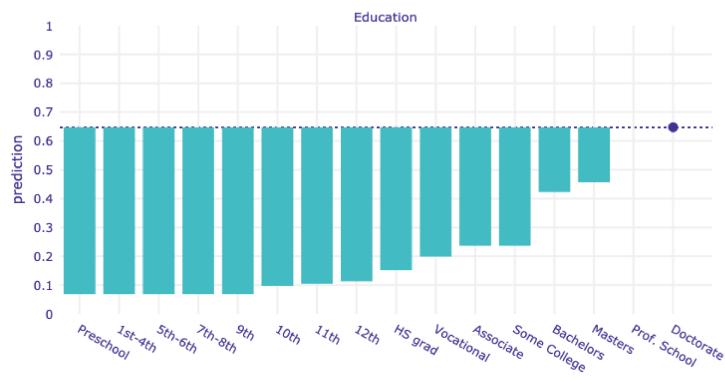
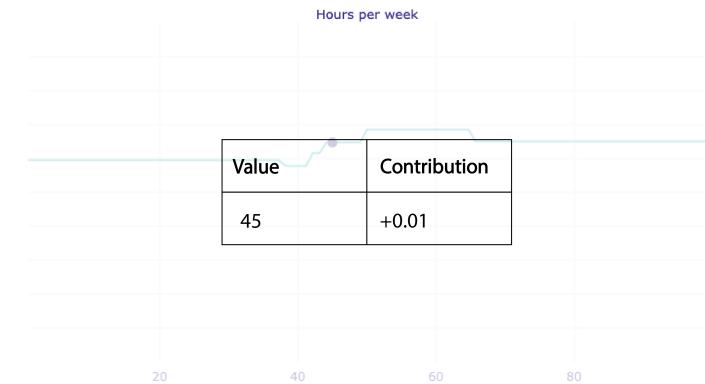
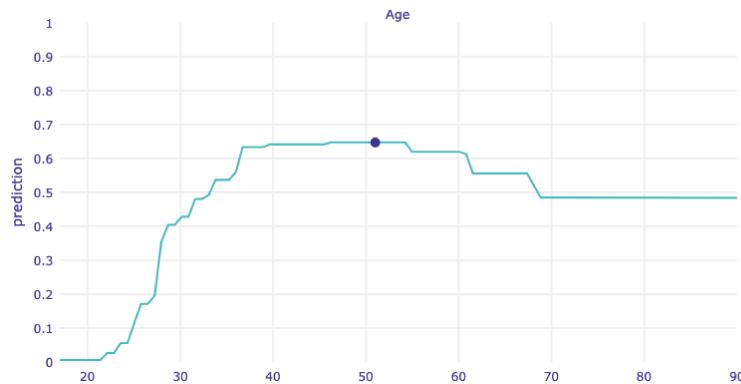
Model Prediction: 0.647



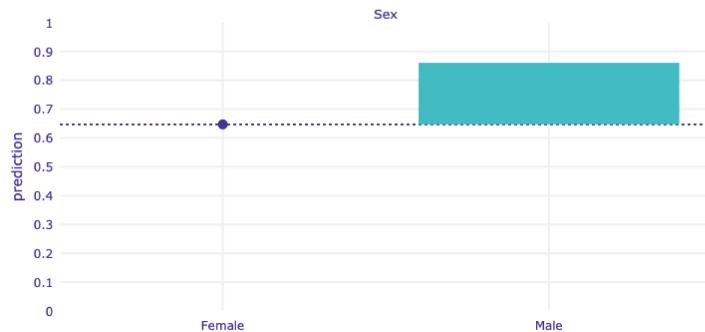
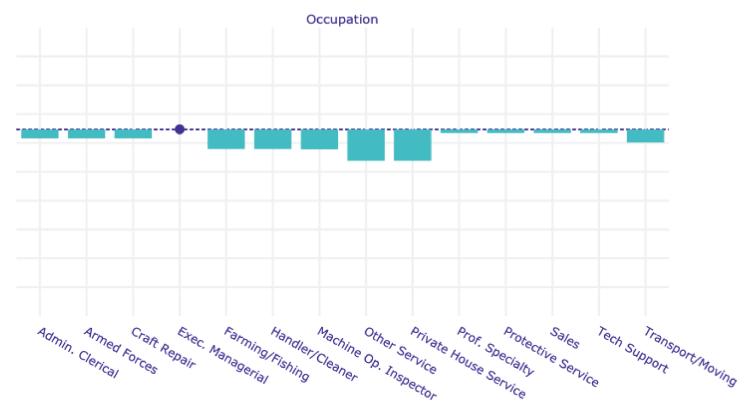
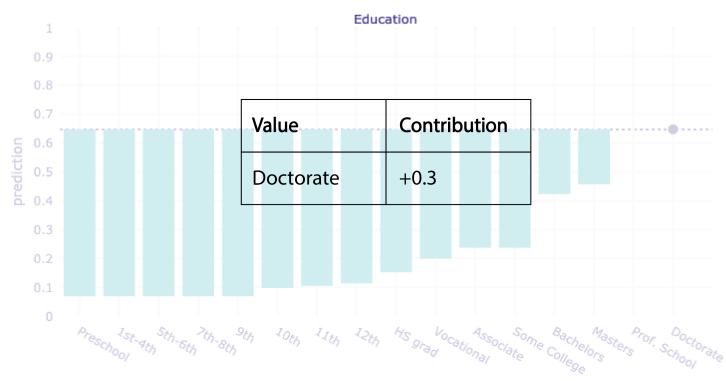
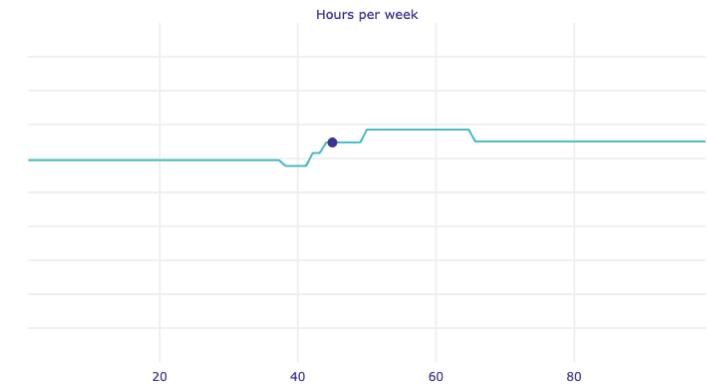
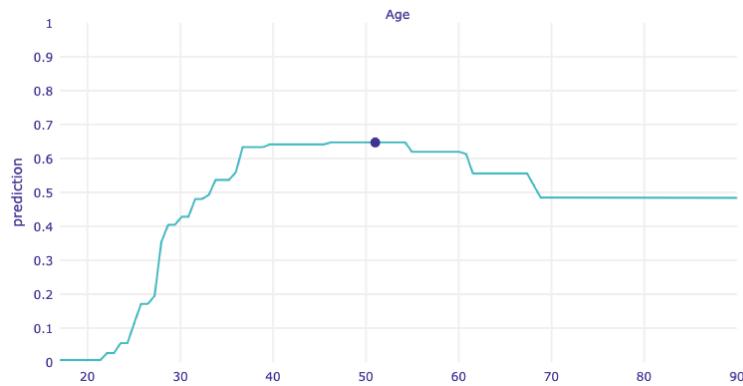
Model Prediction: 0.647



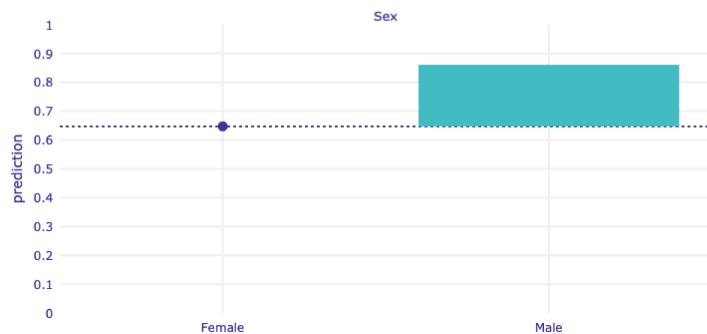
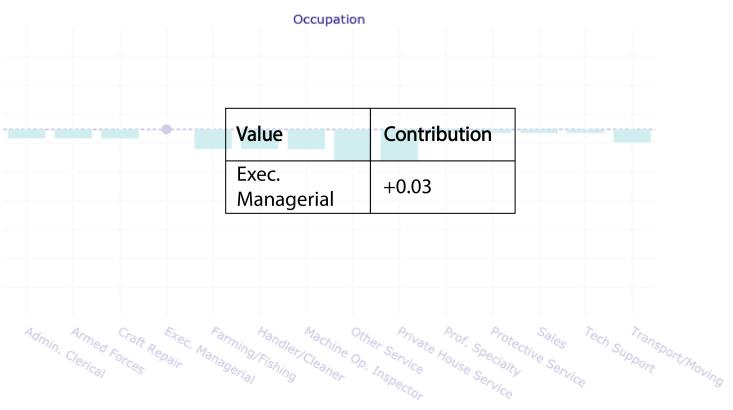
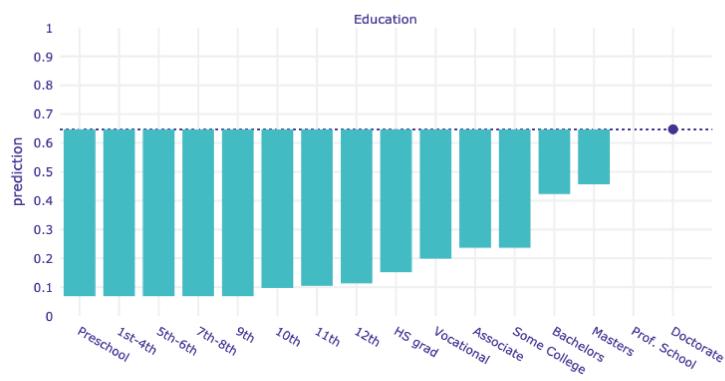
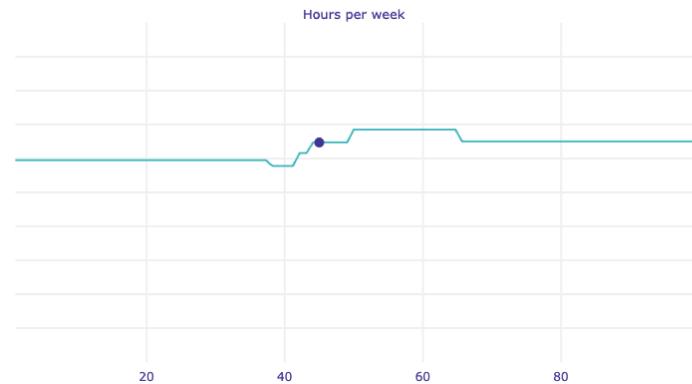
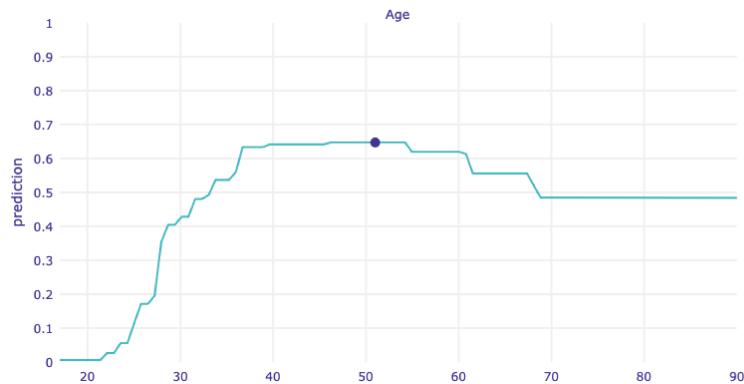
Model Prediction: 0.647



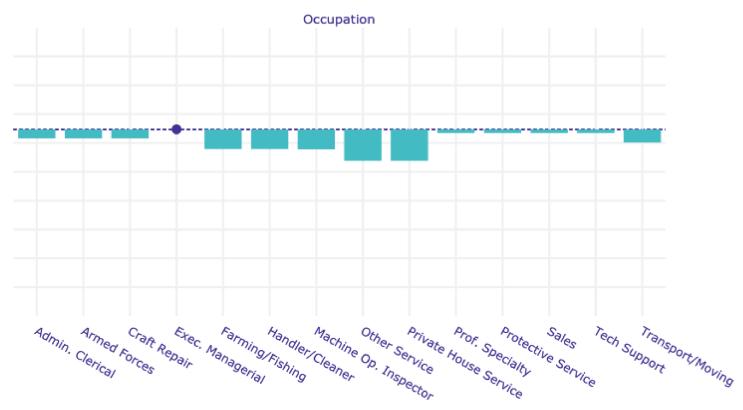
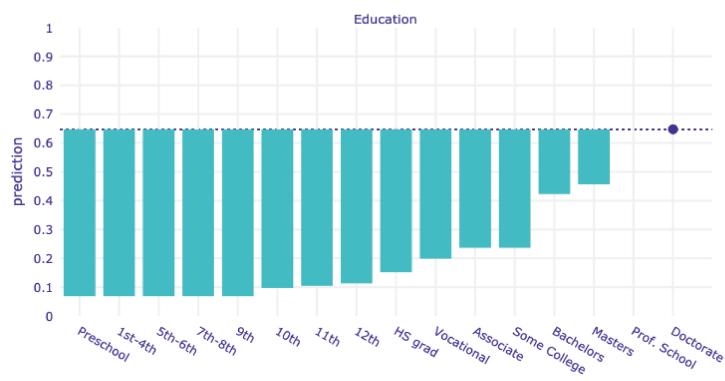
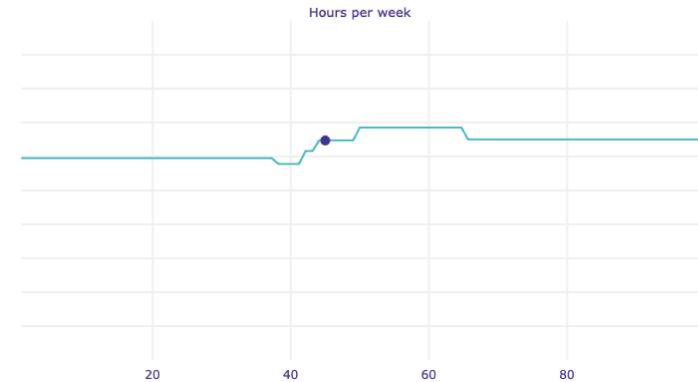
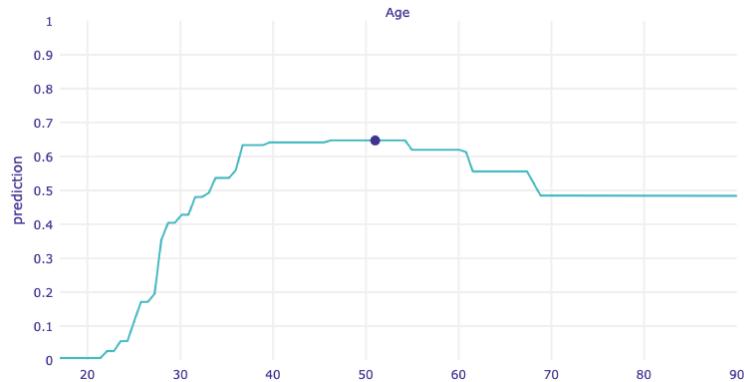
Model Prediction: 0.647



Model Prediction: 0.647



Model Prediction: 0.647



Will this model approve the loan for this person?

YES

NO

Which feature had the most predictive power for this decision?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation

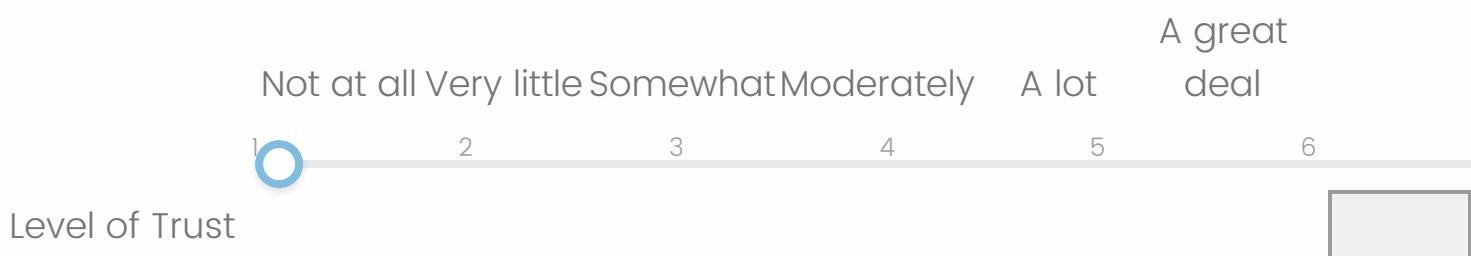
Which factor(s) are pushing the model toward predicting 'NO'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

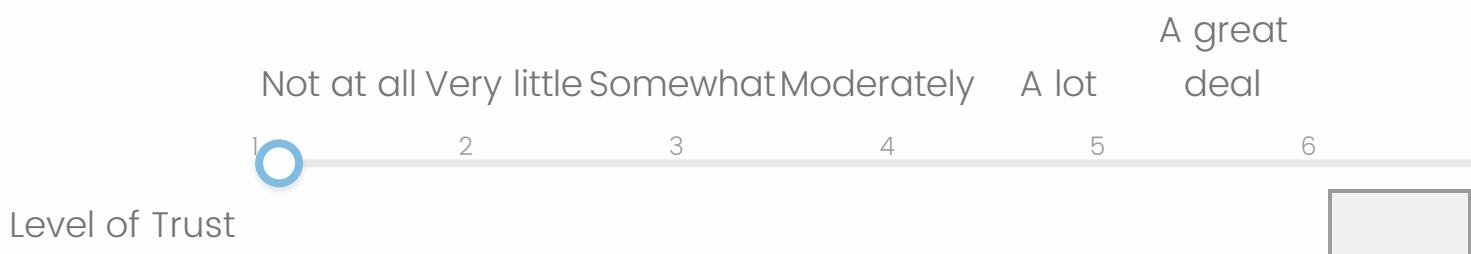
Which factor(s) are pushing the model toward predicting 'YES'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **you**?



On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **other people in general**?



Please indicate whether you agree with the below statements.

Agree

This model uses all of the features that it should use when making this decision.

This model does not use any unnecessary features when making this decision.

I trust the data this model was trained on.

Computer models can be trusted to make human decisions.

This model is accurate.

This model is fair.

This model would probably give me a loan because I am similar to the person described in this question.

This model would probably give me a loan because I am different from the person described in this question.

This model would probably give me a loan because of previous decisions it has made.

This model probably would not give me a loan, and this would be the correct decision.

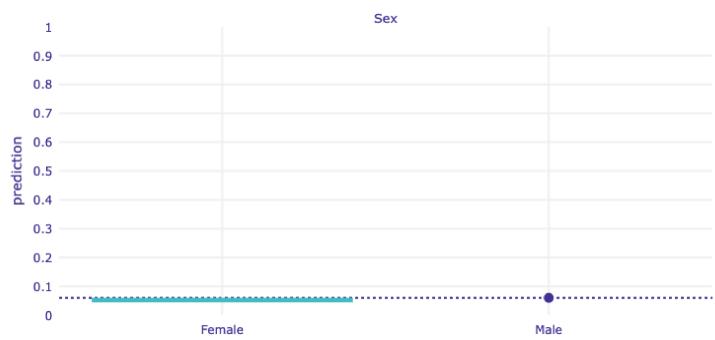
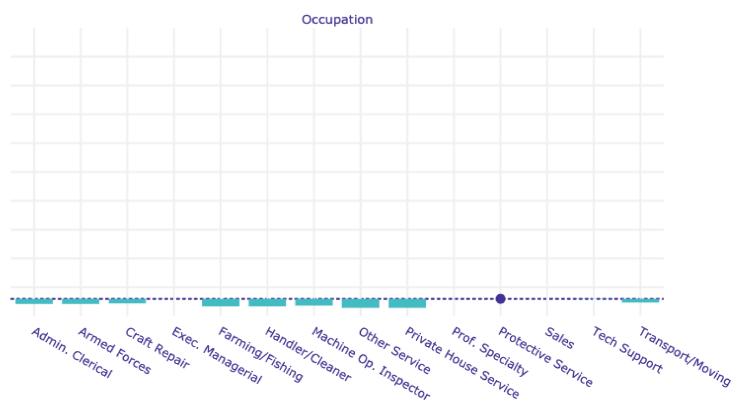
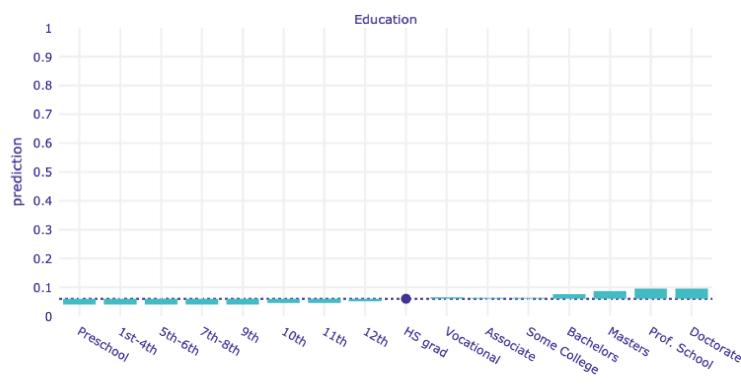
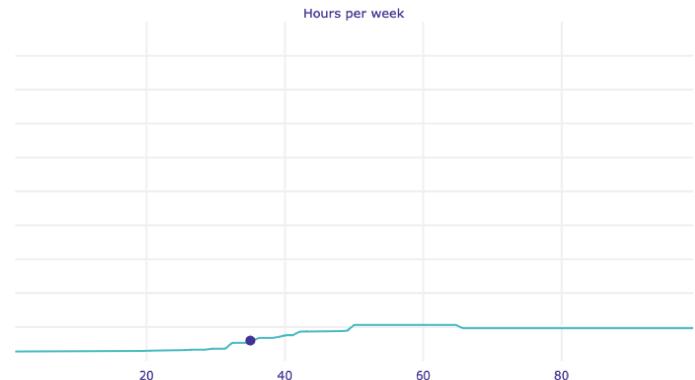
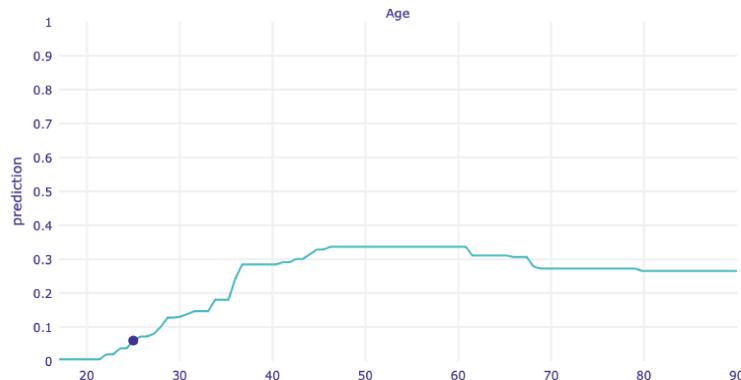
Man 1

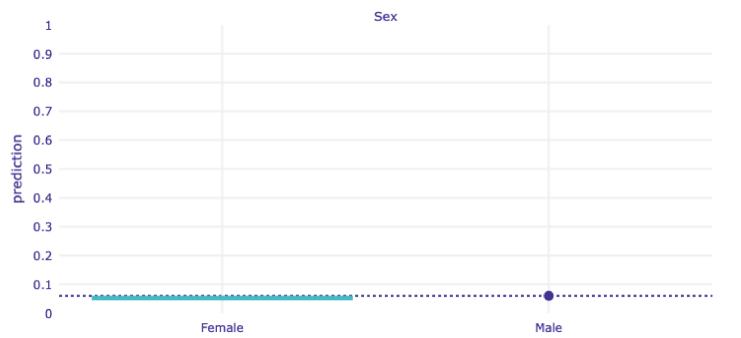
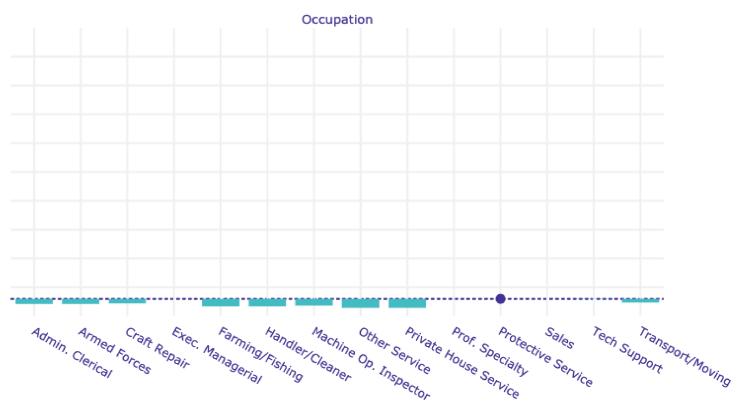
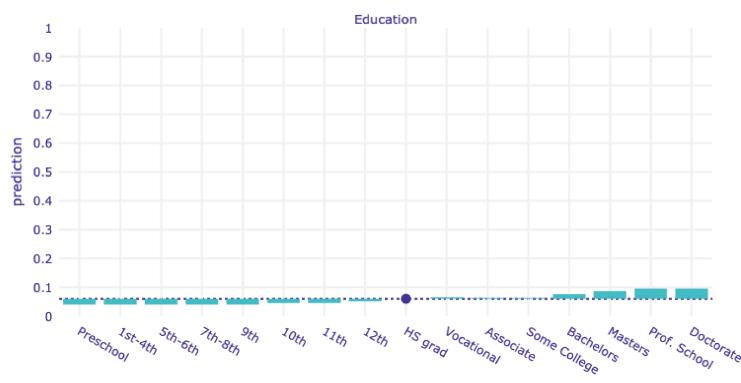
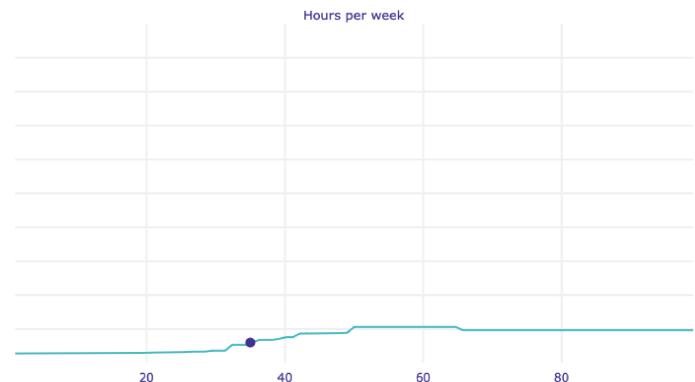
Below you will find the information of Applicant N.

You can see that the model made a prediction of whether to approve or deny a loan from this applicant based on five factors. The explanation is below.

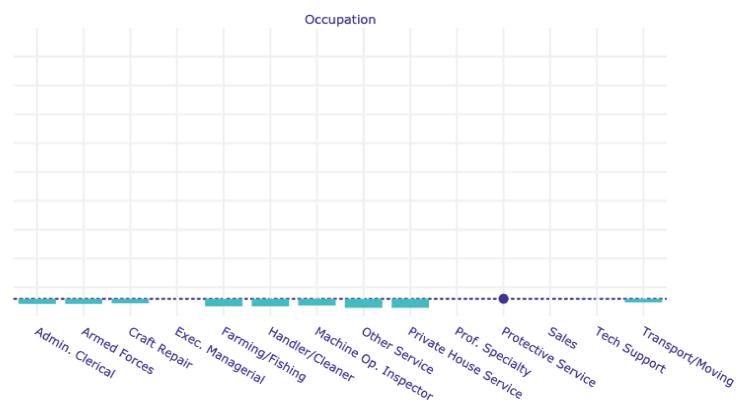
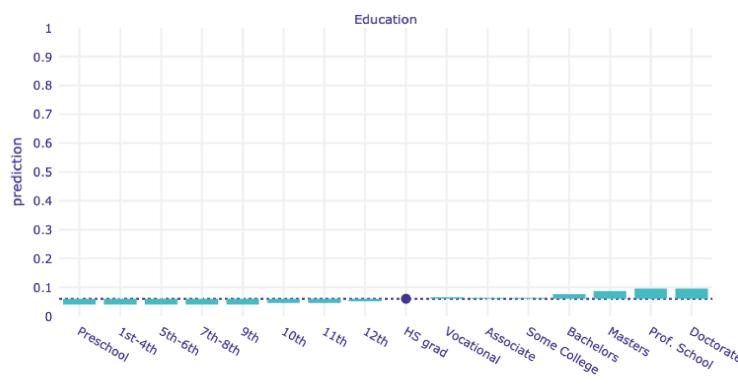
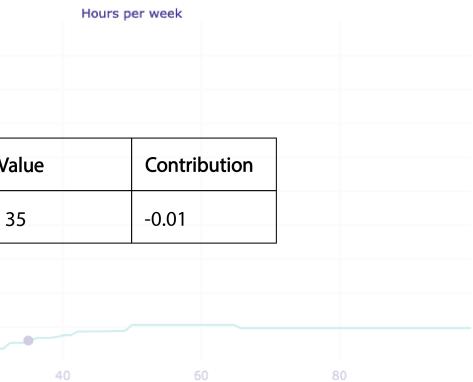
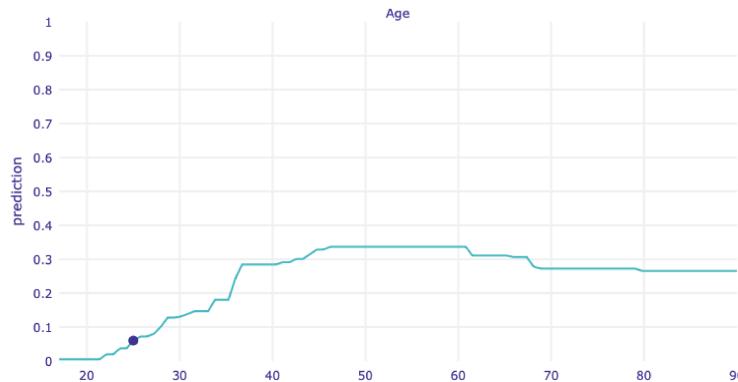
Look at the explanation, and answer the questions that follow.

Remember that if the model's prediction probability (Predicted Value) for 'YES' is greater than or equal to 0.5, the model will return 'YES' (approve the loan). If it is less than 0.5, the model will return 'NO' (deny the loan).

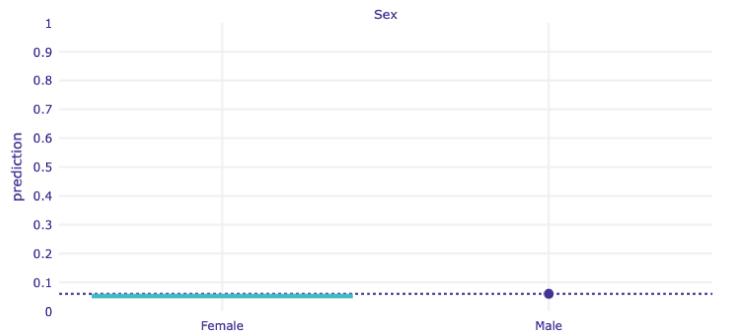
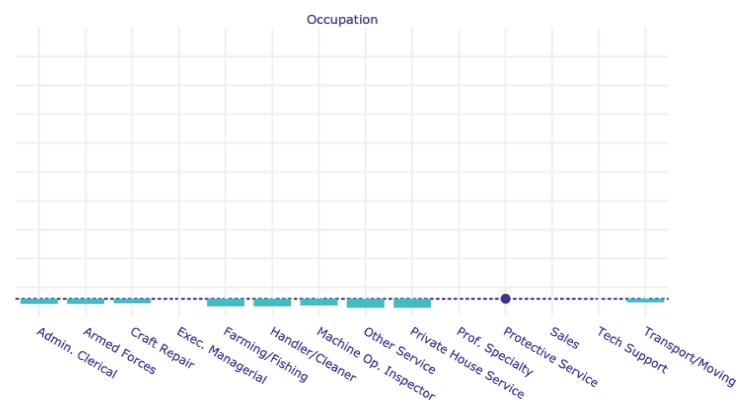
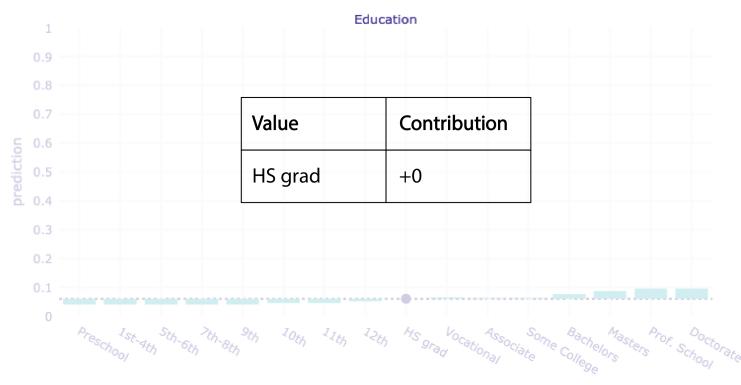
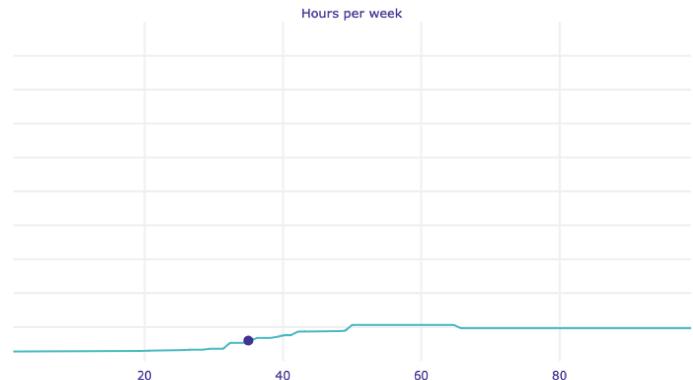
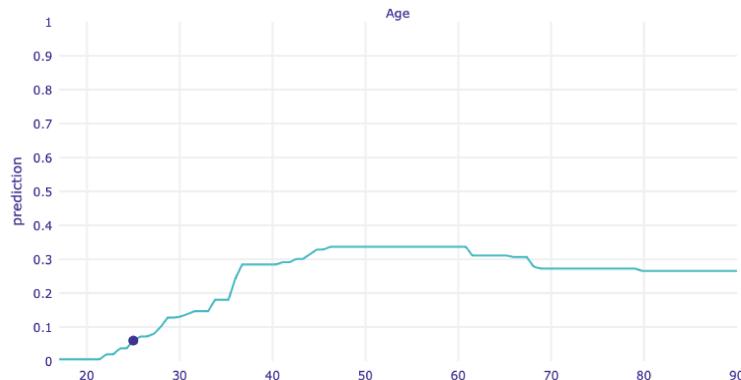
Model Prediction: 0.06

Model Prediction: 0.06

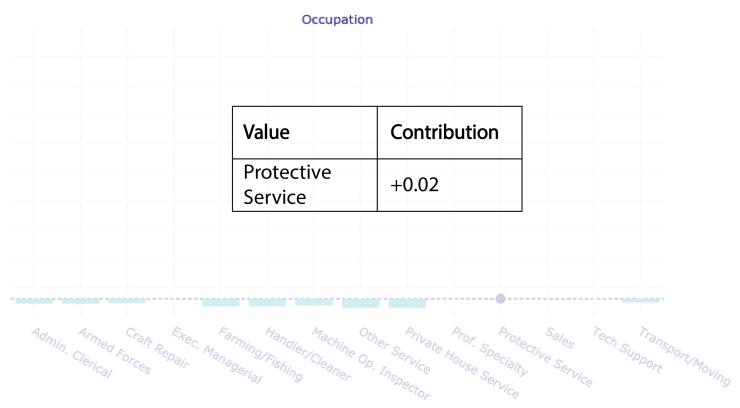
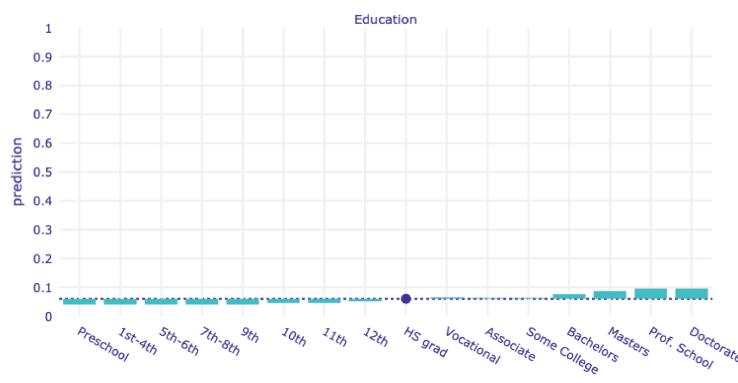
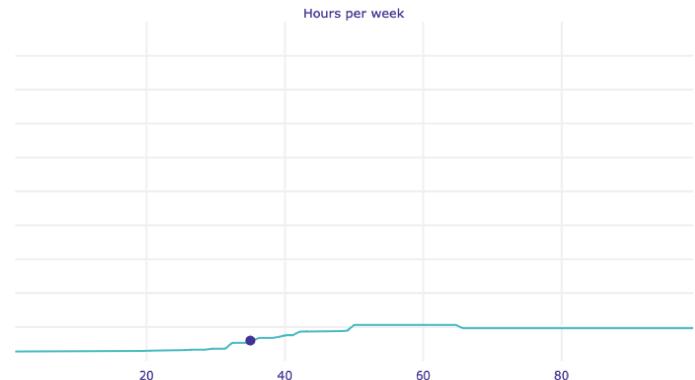
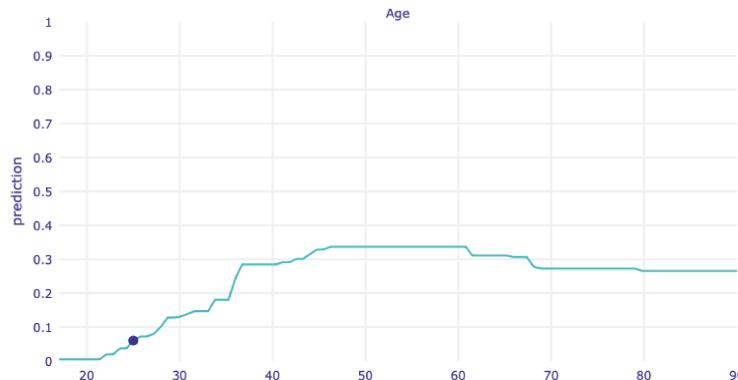
Model Prediction: 0.06



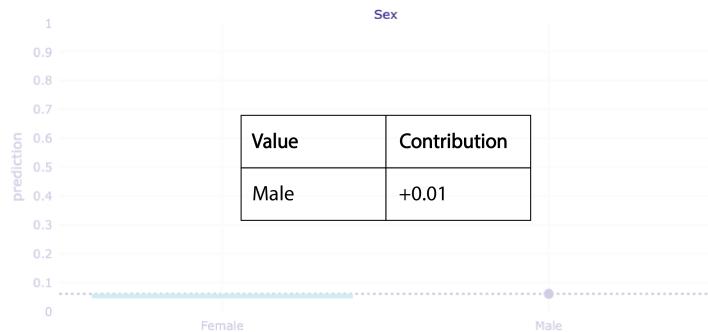
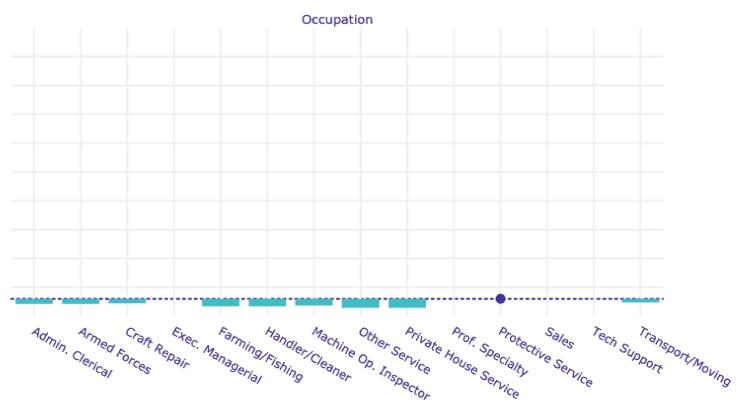
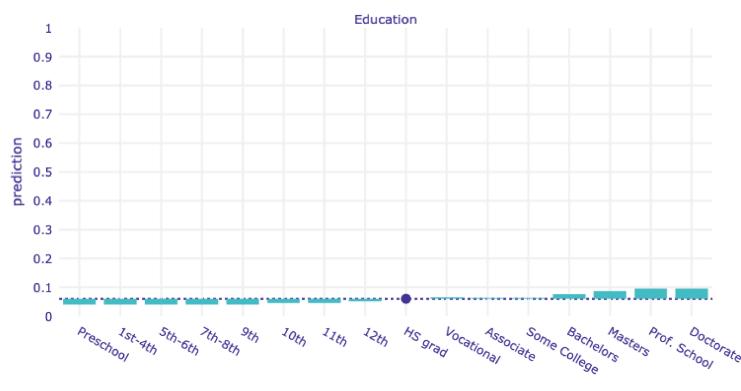
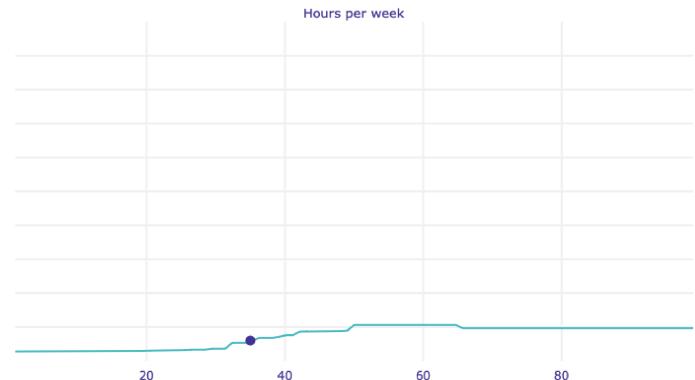
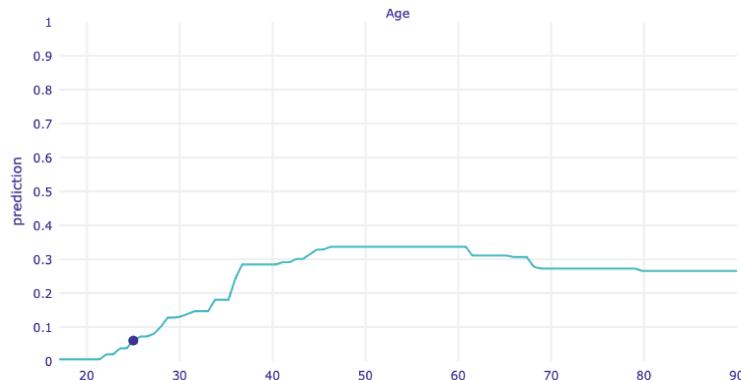
Model Prediction: 0.06



Model Prediction: 0.06



Model Prediction: 0.06



Value	Contribution
Male	+0.01

Will this model approve the loan for this person?

YES

NO

Which feature had the most predictive power for this decision?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation

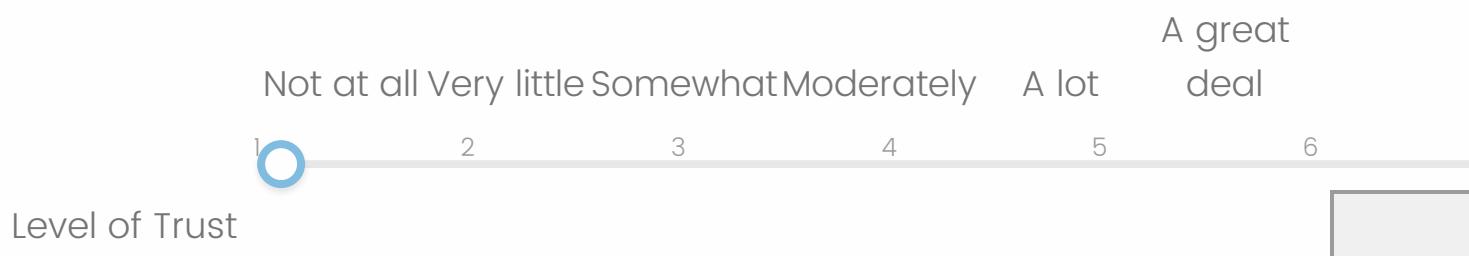
Which factor(s) are pushing the model toward predicting 'NO'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

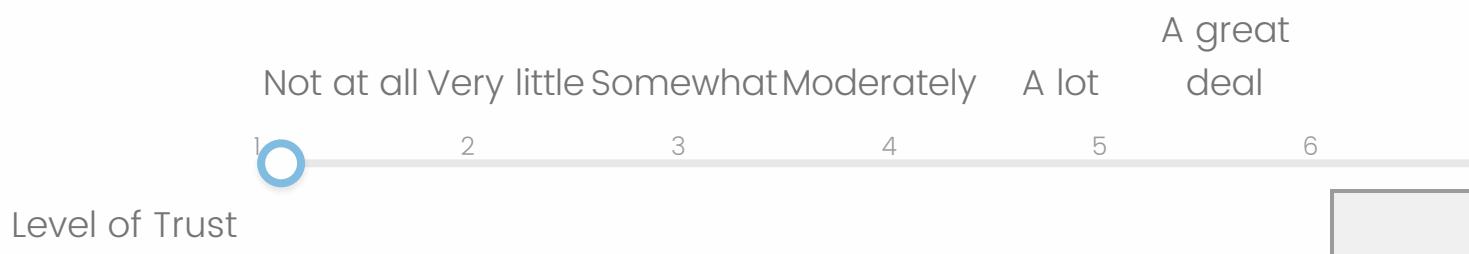
Which factor(s) are pushing the model toward predicting 'YES'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **you**?



On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **other people in general**?



Please indicate whether you agree with the below statements.

Agree

This model uses all of the features that it should use when making this decision.

This model does not use any unnecessary features when making this decision.

I trust the data this model was trained on.

Computer models can be trusted to make human decisions.

This model is accurate.

This model is fair.

This model would probably give me a loan because I am similar to the person described in this question.

This model would probably give me a loan because I am different from the person described in this question.

This model would probably give me a loan because of previous decisions it has made.

This model probably would not give me a loan, and this would be the correct decision.

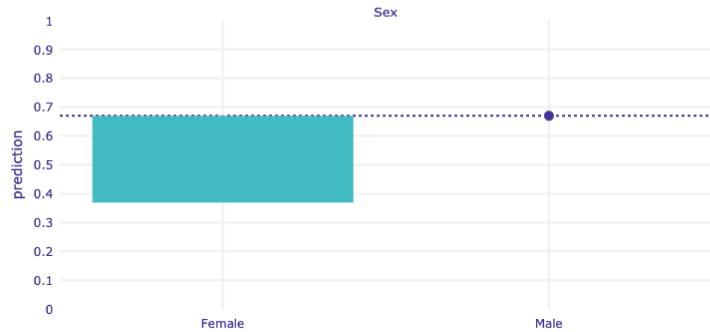
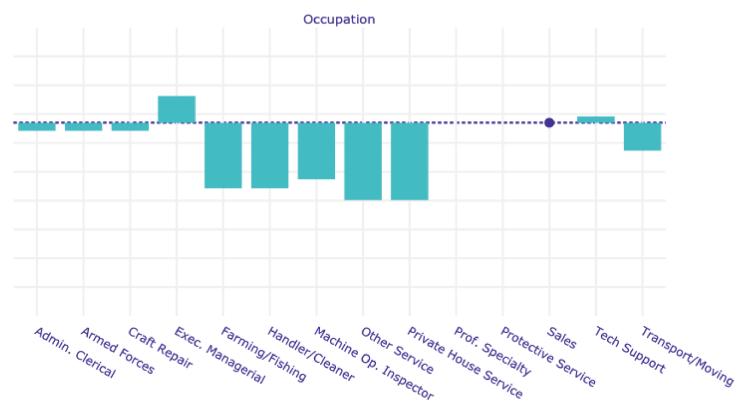
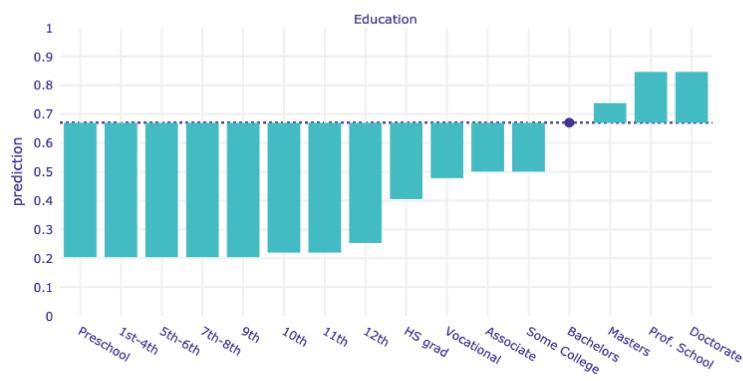
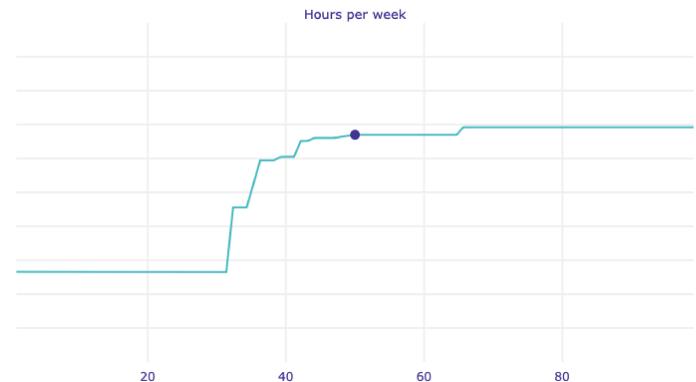
Man 2

Below you will find the information of Applicant P.

You can see that the model made a prediction of whether to approve or deny a loan from this applicant based on five factors.

The explanation is below. Look at the explanation, and answer the questions that follow.

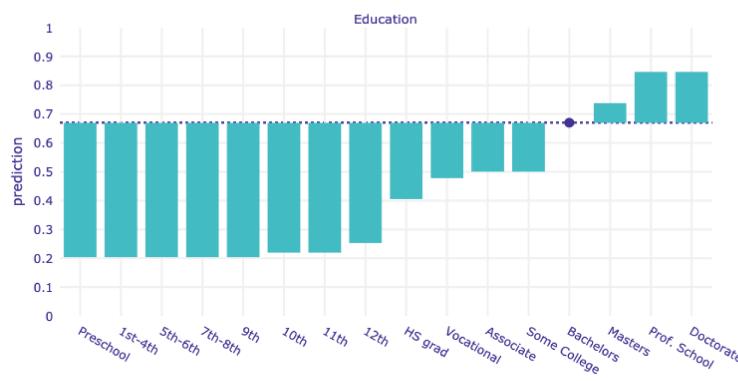
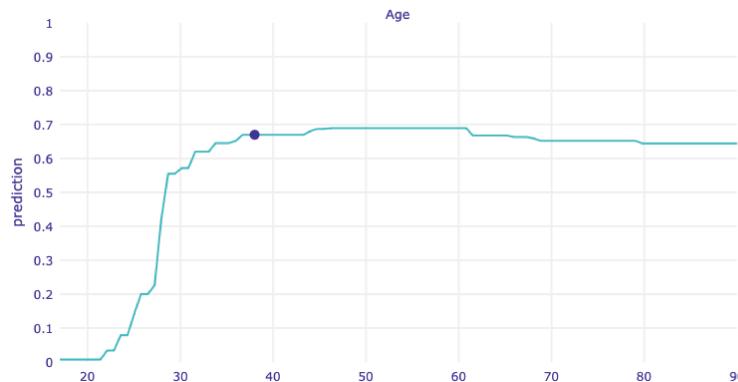
Remember that if the model's prediction probability (Predicted Value) for 'YES' is greater than or equal to 0.5, the model will return 'YES' (approve the loan). If it is less than 0.5, the model will return 'NO' (deny the loan).

Model Prediction: 0.67

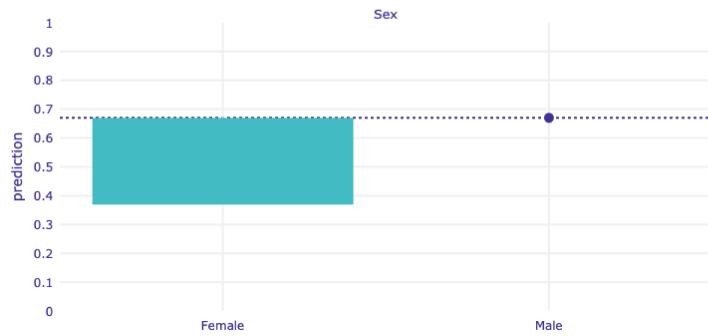
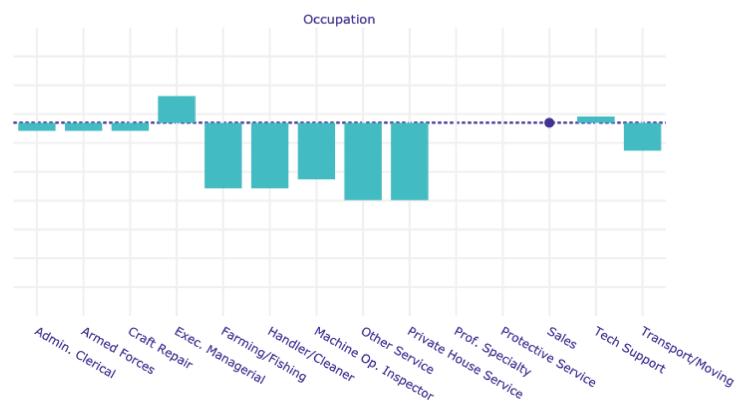
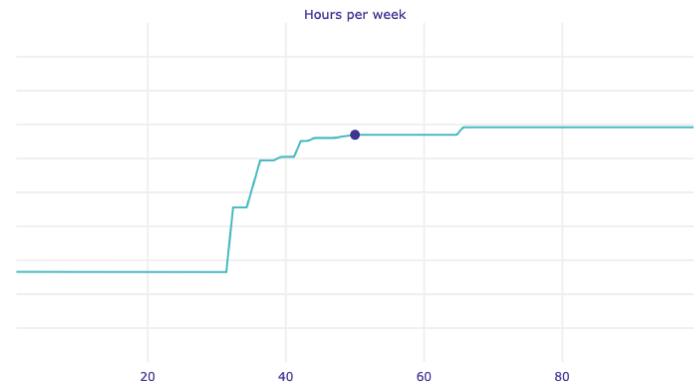
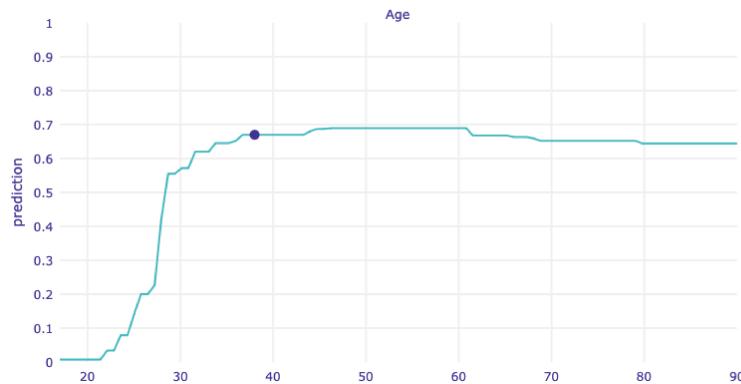
Model Prediction: 0.67



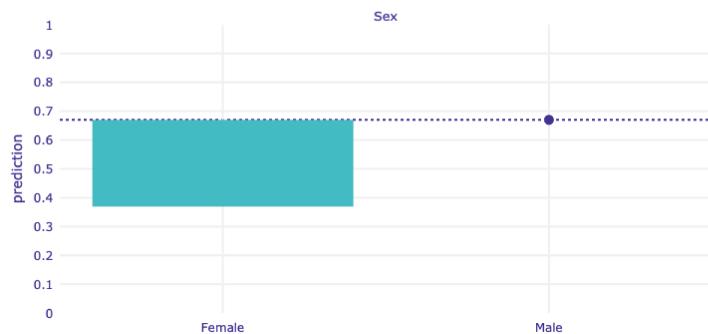
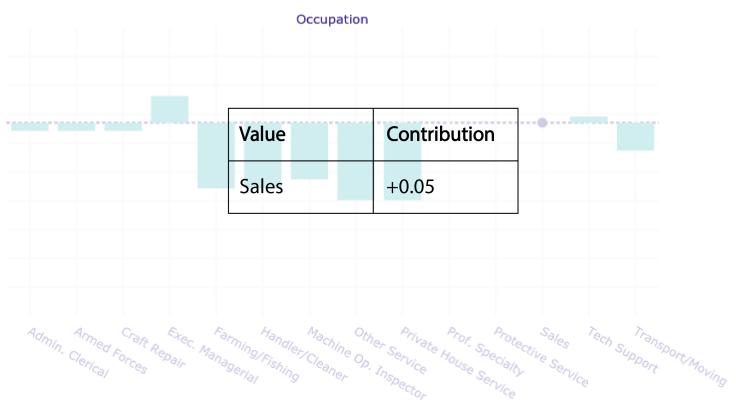
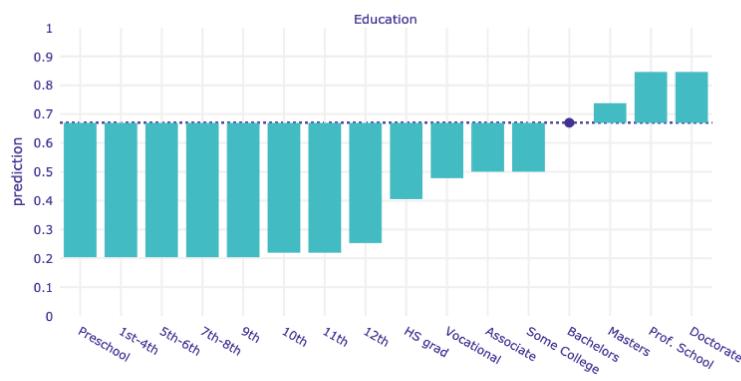
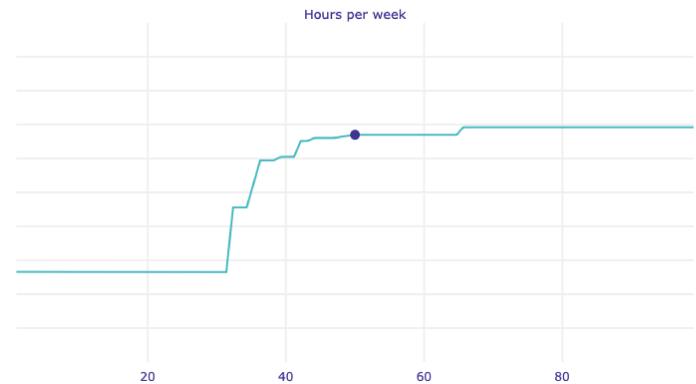
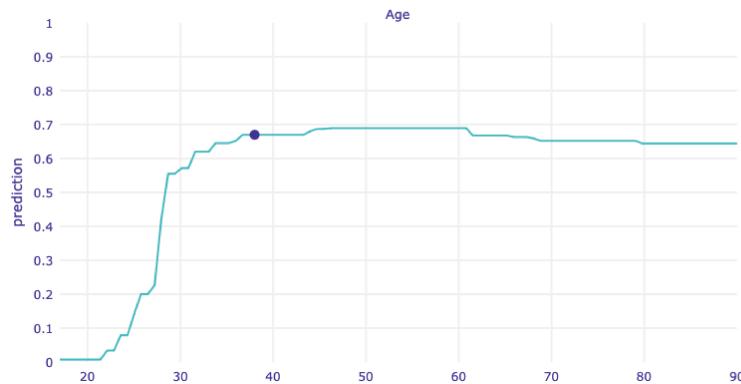
Model Prediction: 0.67



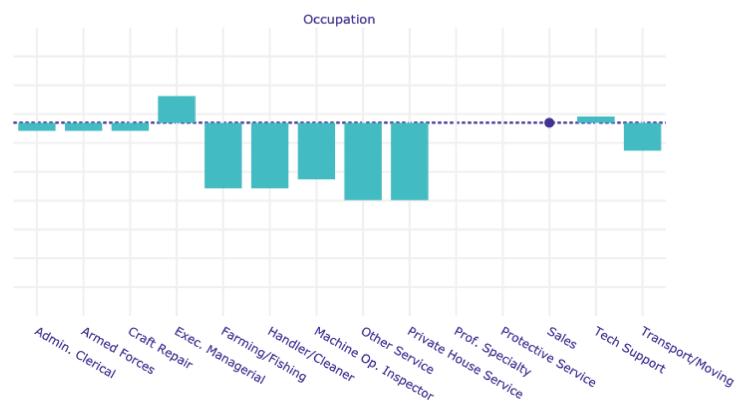
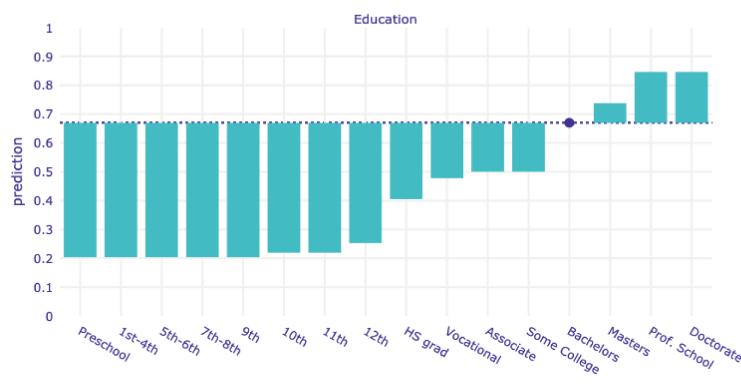
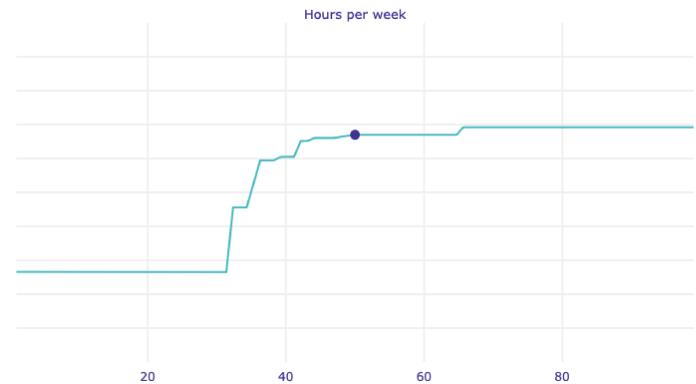
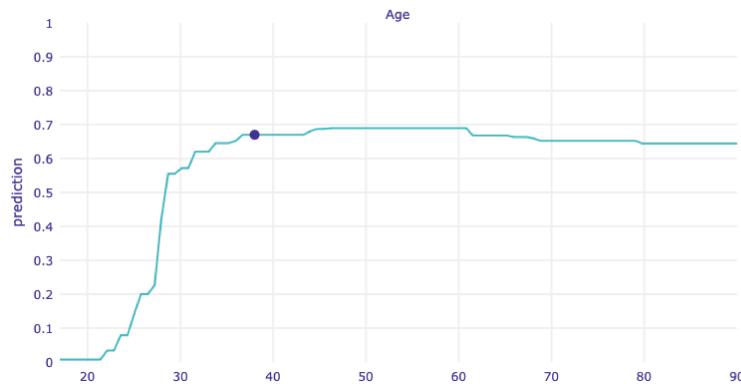
Model Prediction: 0.67



Model Prediction: 0.67



Model Prediction: 0.67



Will this model approve the loan for this person?

YES

NO

Which feature had the most predictive power for this decision?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation

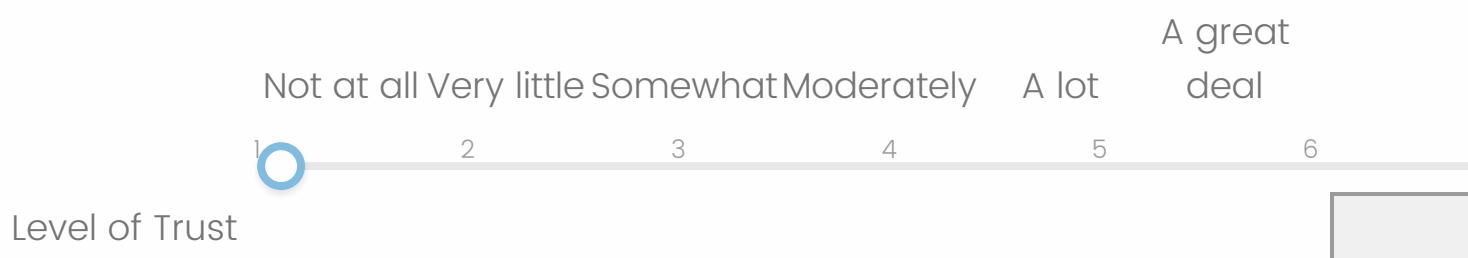
Which factor(s) are pushing the model toward predicting 'NO'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

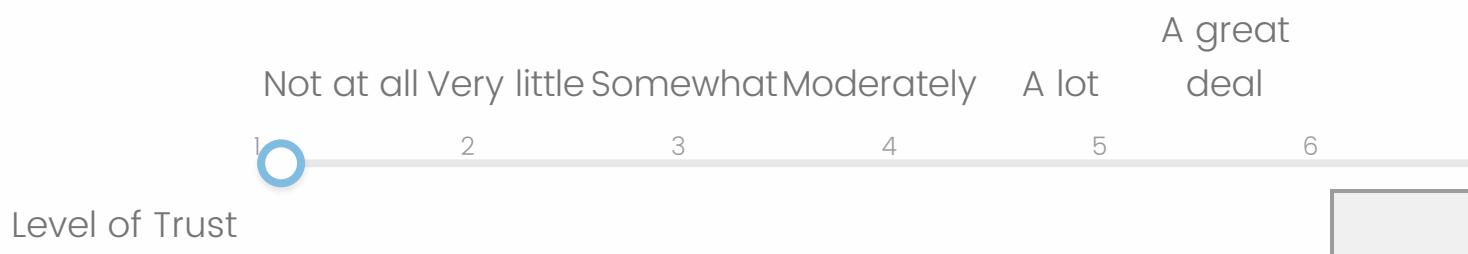
Which factor(s) are pushing the model toward predicting 'YES'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **you**?



On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **other people in general**?



Please indicate whether you agree with the below statements.

Agree

This model uses all of the features that it should use when making this decision.

This model does not use any unnecessary features when making this decision.

I trust the data this model was trained on.

Computer models can be trusted to make human decisions.

This model is accurate.

This model is fair.

This model would probably give me a loan because I am similar to the person described in this question.

This model would probably give me a loan because I am different from the person described in this question.

This model would probably give me a loan because of previous decisions it has made.

This model probably would not give me a loan, and this would be the correct decision.

Man 3

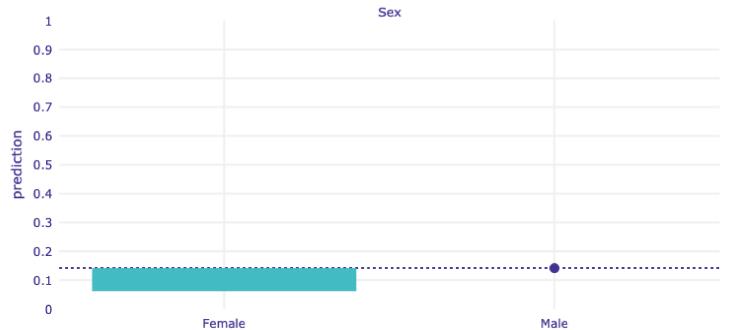
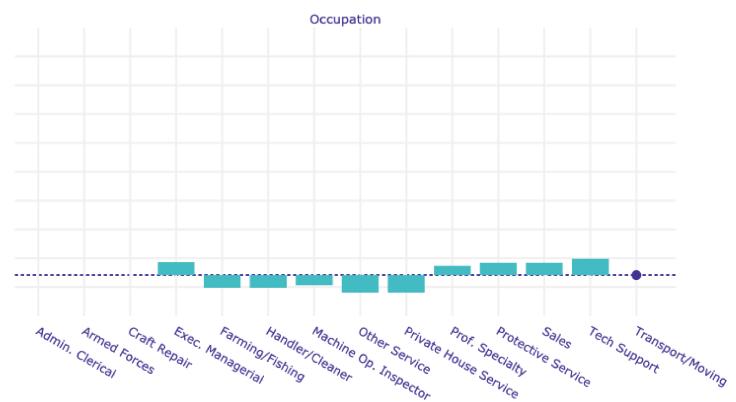
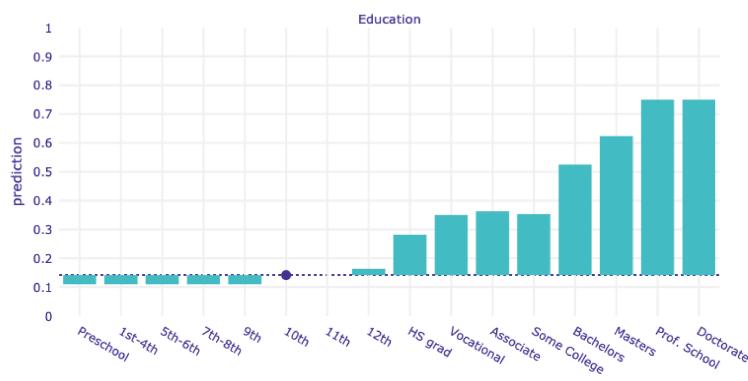
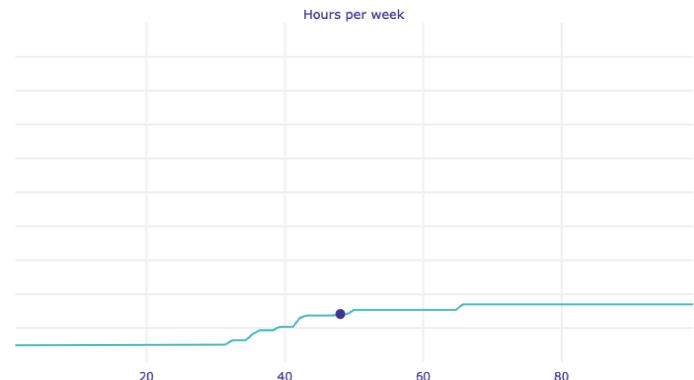
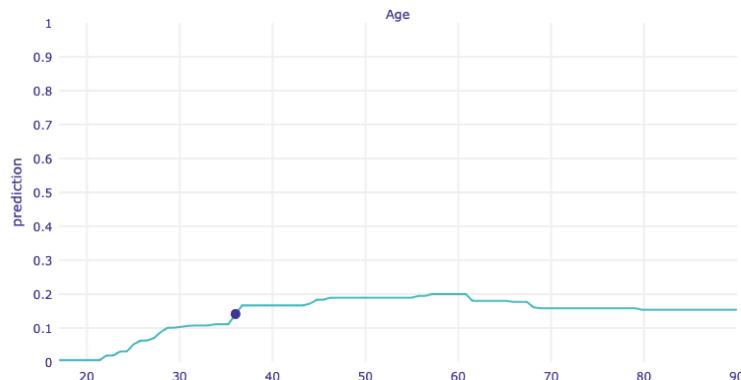
Below you will find the information of Applicant K.

You can see that the model made a prediction of whether to approve or deny a loan from this applicant based on five factors. The explanation is below.

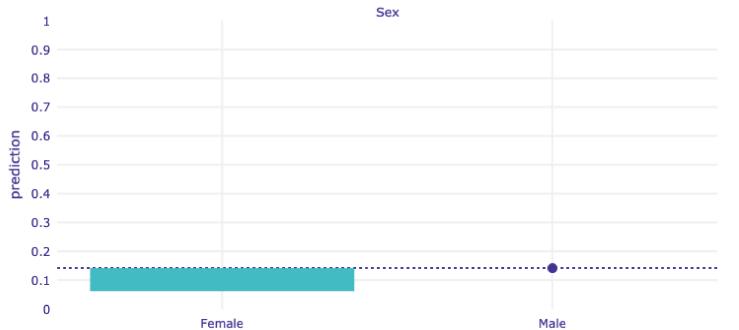
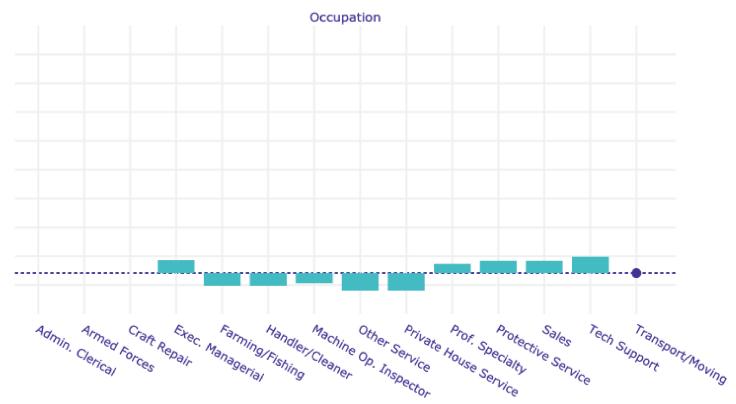
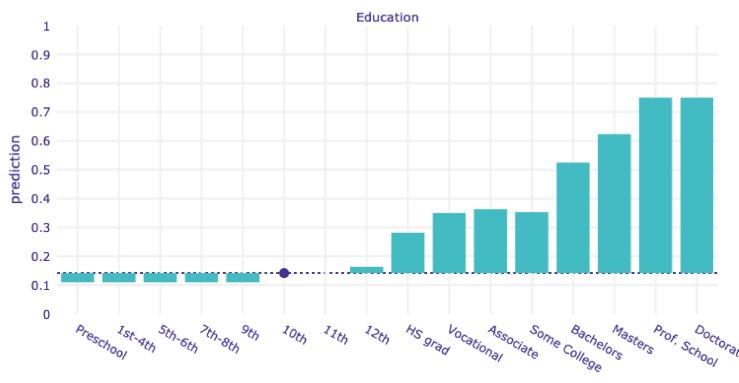
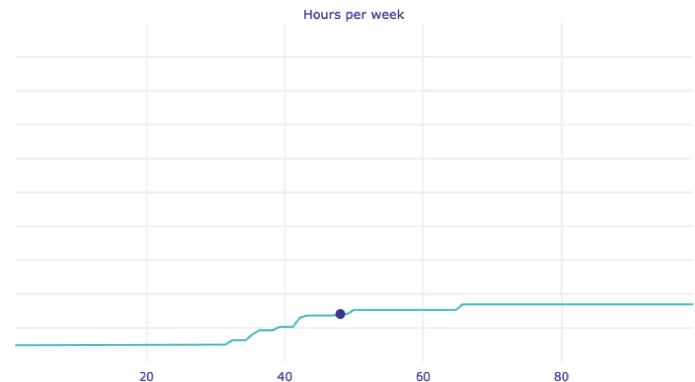
Look at the explanation, and answer the questions that follow.

Remember that if the model's prediction probability (Predicted Value) for 'YES' is greater than or equal to 0.5, the model will return 'YES' (approve the loan). If it is less than 0.5, the model will return 'NO' (deny the loan).

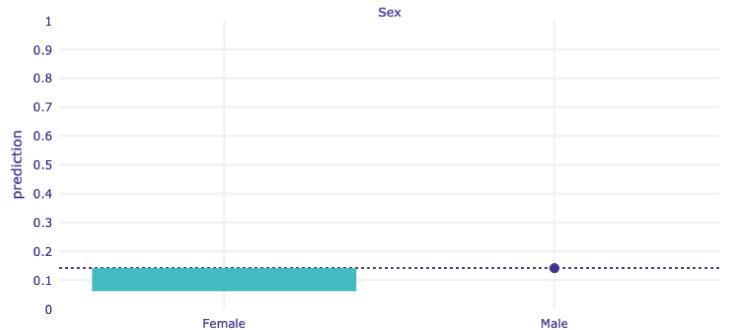
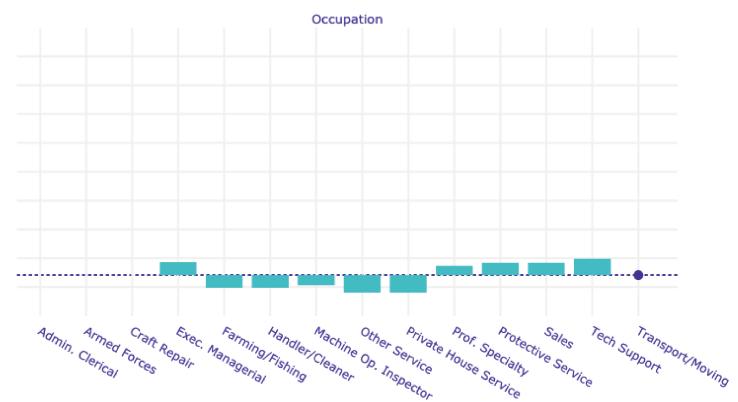
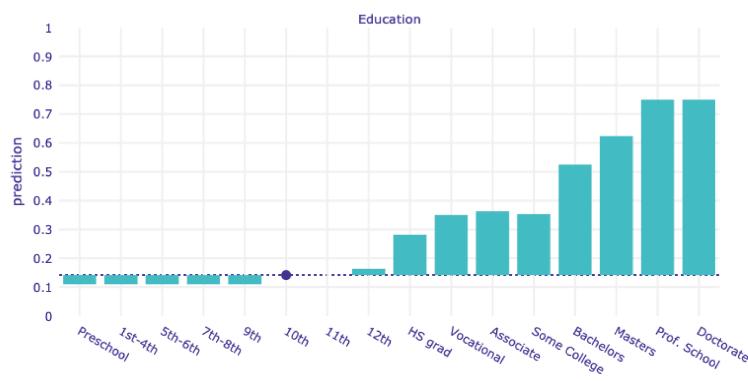
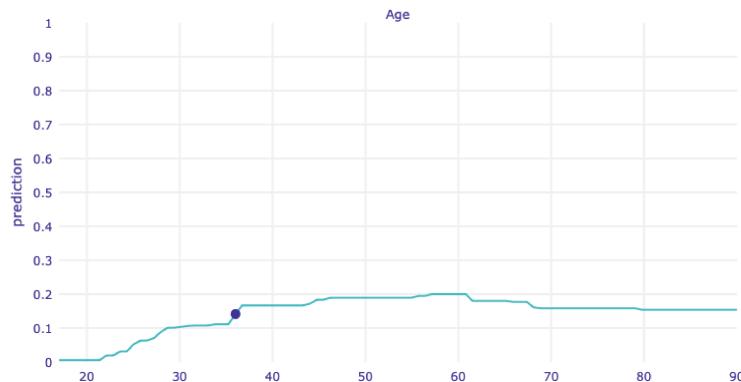
Model Prediction: 0.142



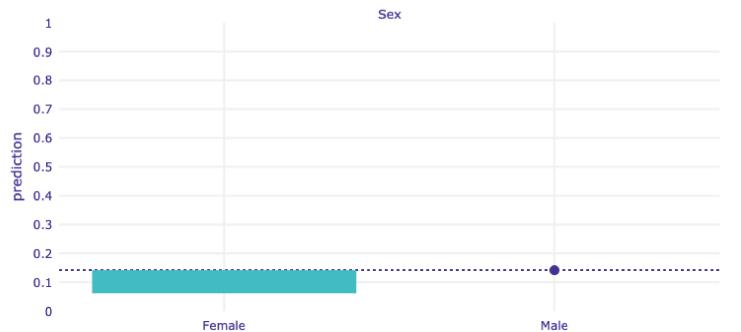
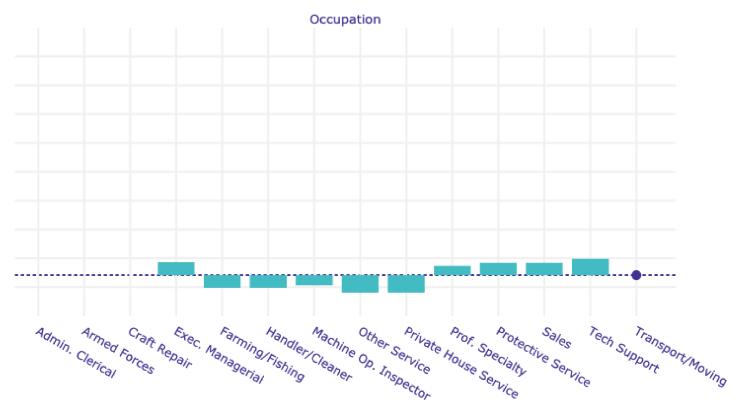
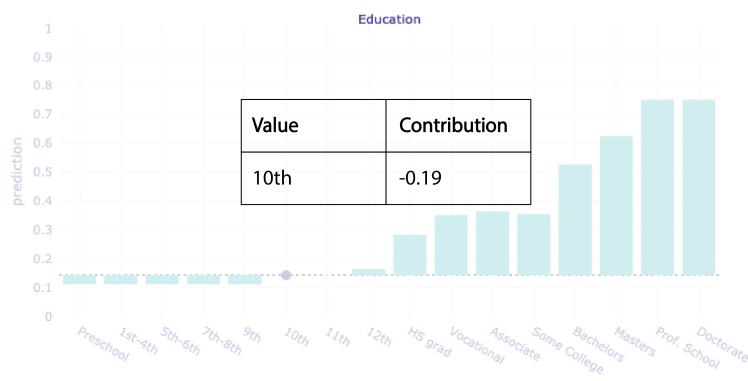
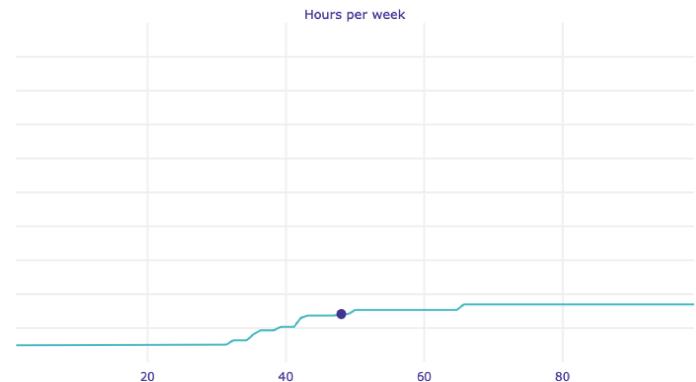
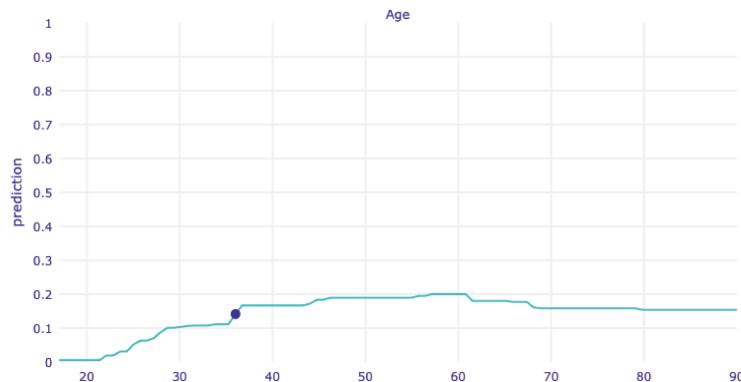
Model Prediction: 0.142



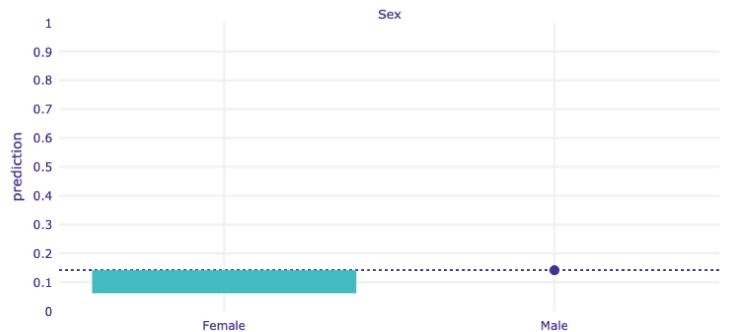
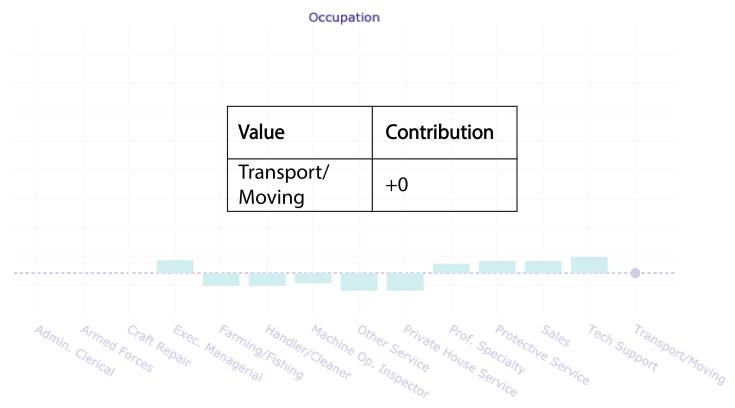
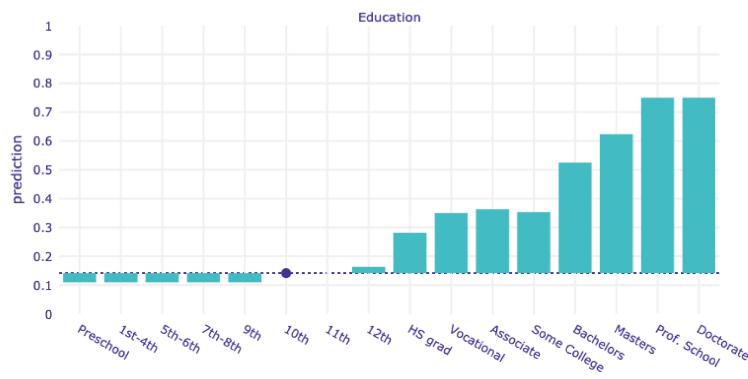
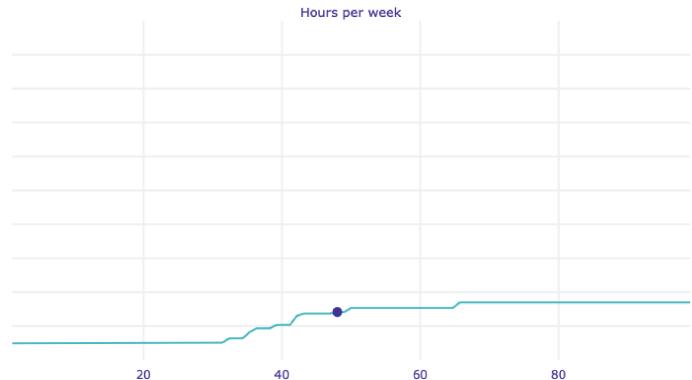
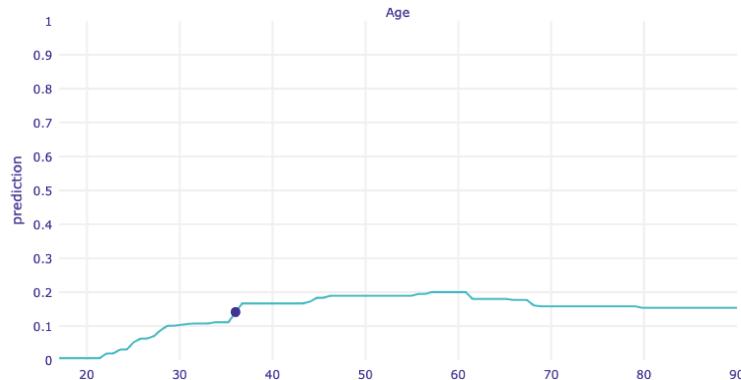
Model Prediction: 0.142



Model Prediction: 0.142

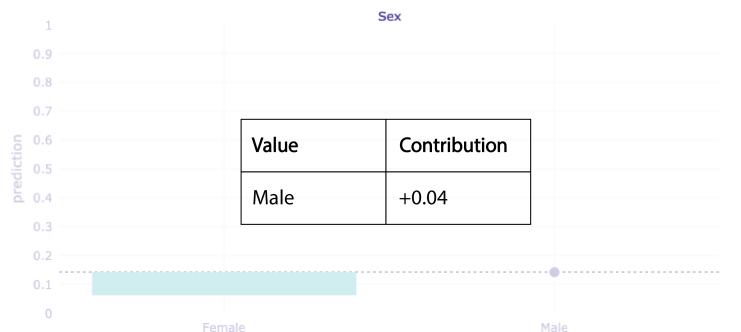
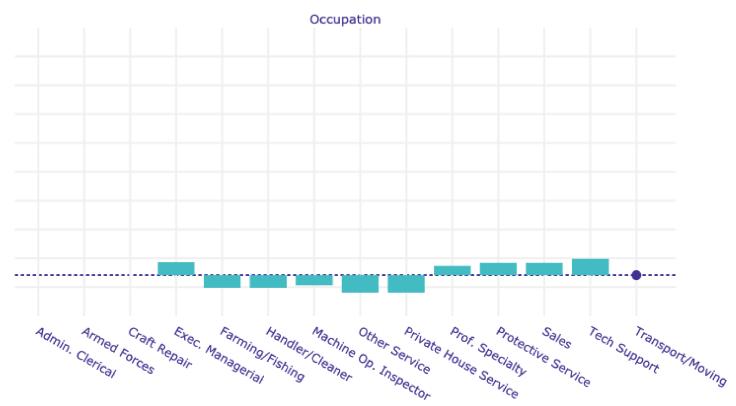
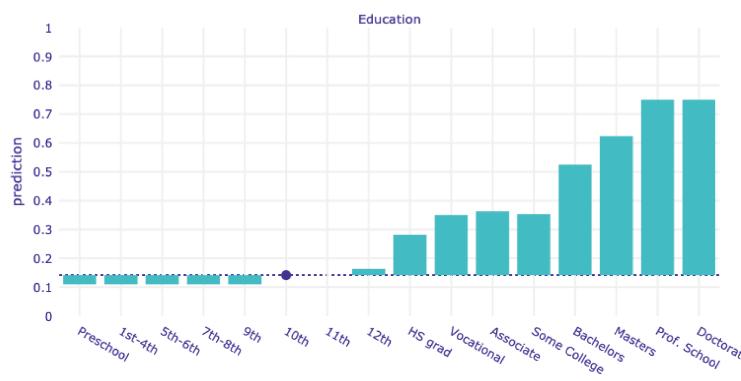
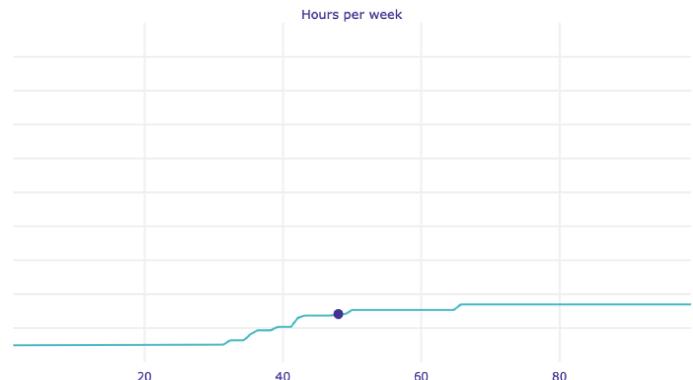
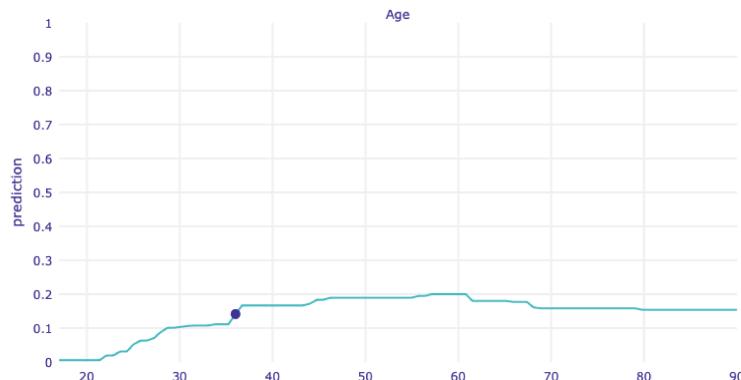


Model Prediction: 0.142



Value	Contribution
Transport/ Moving	+0

Model Prediction: 0.142



Will this model approve the loan for this person?

YES

NO

What feature had the most predictive power for this decision?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation

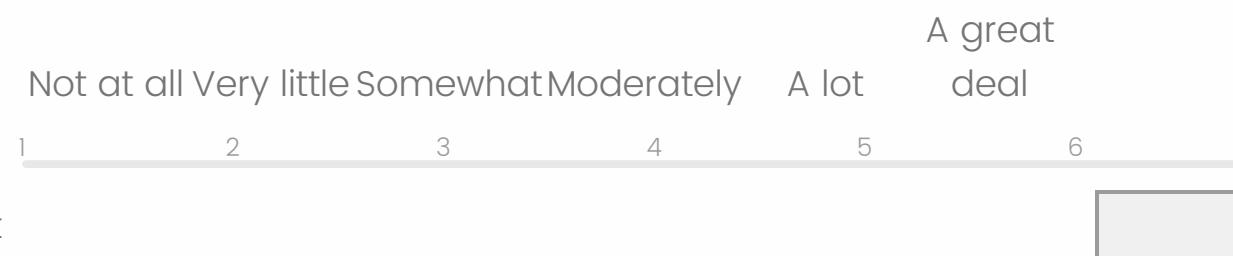
Which factor(s) are pushing the model toward predicting 'NO'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

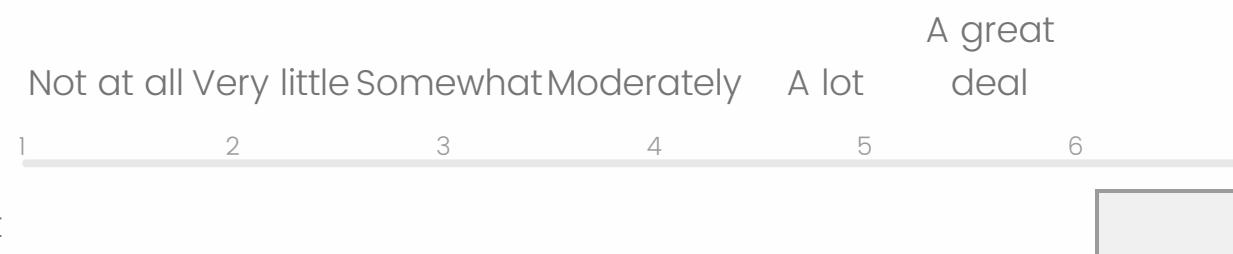
Which factor(s) are pushing the model toward predicting 'YES'?

- Education
- Hours Worked Per Week
- Age
- Sex
- Occupation
- None of these

On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **you**?



On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **other people in general**?



Please indicate whether you agree with the below statements.

Agree

This model uses all of the features that it should use when making this decision.

This model does not use any unnecessary features when making this decision.

I trust the data this model was trained on.

Computer models can be trusted to make human decisions.

This model is accurate.

This model is fair.

This model would probably give me a loan because I am similar to the person described in this question.

This model would probably give me a loan because I am different from the person described in this question.

This model would probably give me a loan because of previous decisions it has made.

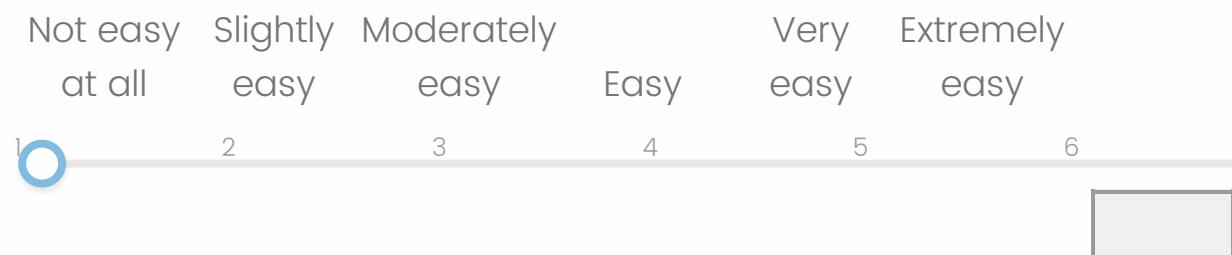
This model probably would not give me a loan, and this would be the correct decision.

Perception of understanding

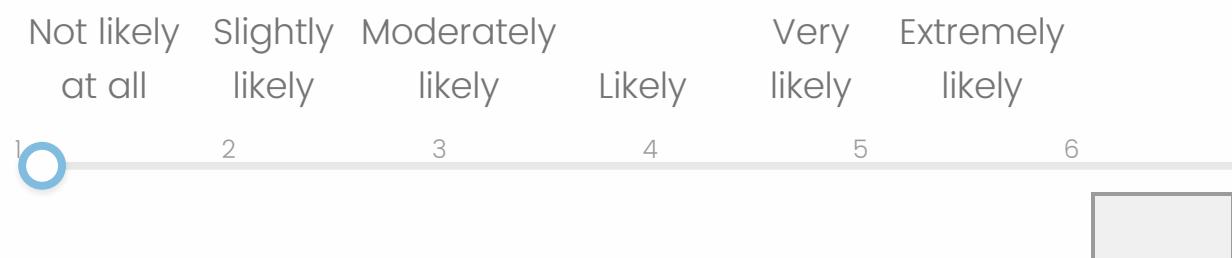
How well did you understand the way this model makes decisions?



How easy was it for you to understand the model output?



How likely would you use this visualization to explain models to other people?



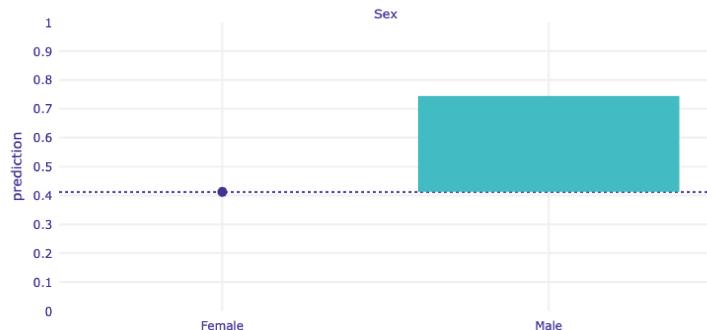
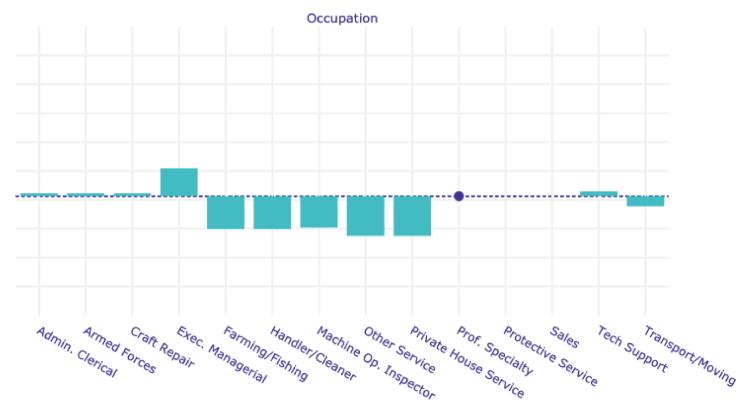
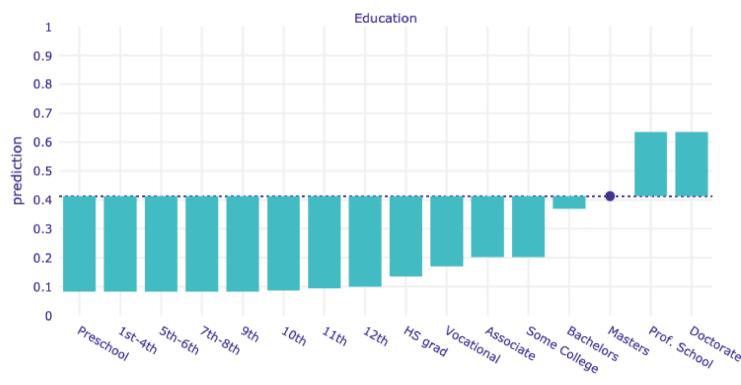
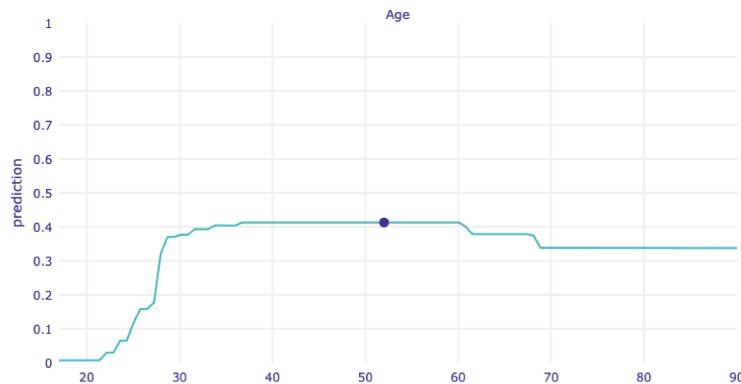
Fairness

Below are two explanations for predictions made by the same loan approval machine learning model you have been seeing, for two people with almost identical features.

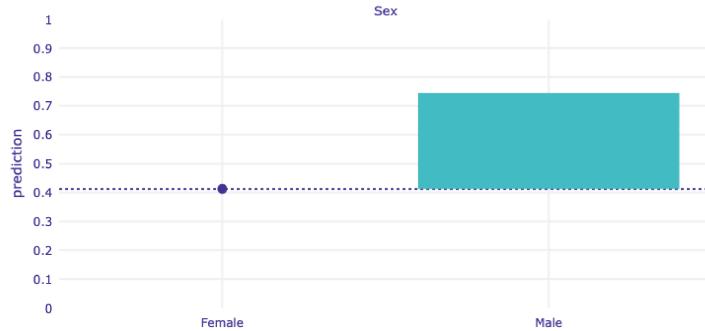
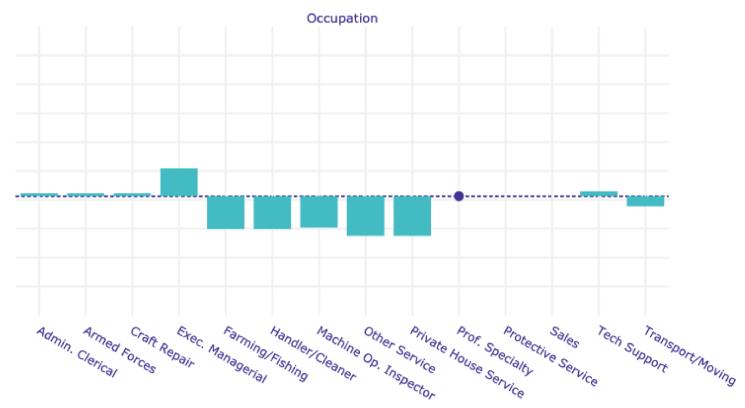
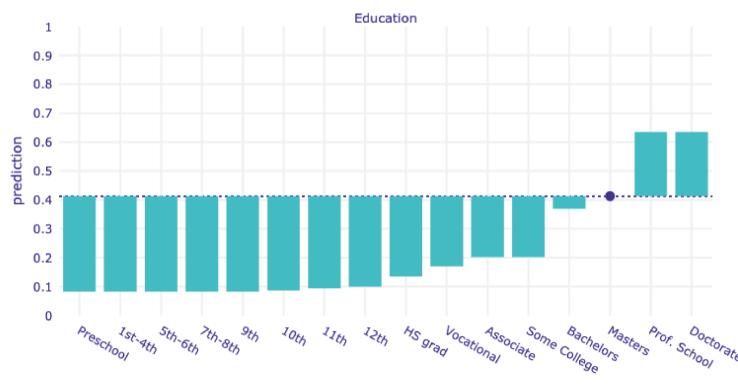
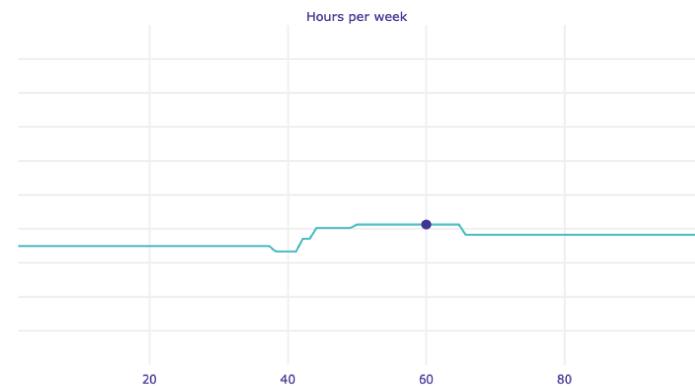
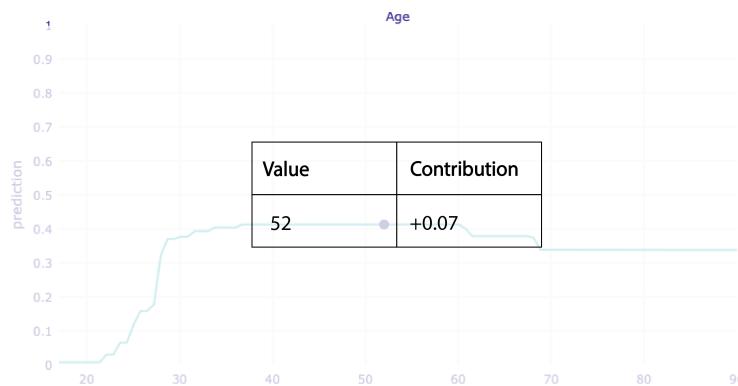
Remember that if the model's prediction probability (Predicted Value) for 'YES' is greater than or equal to 0.5, the model will return 'YES' (approve the loan). If it is less than 0.5, the model will return 'NO' (deny the loan).

Person A

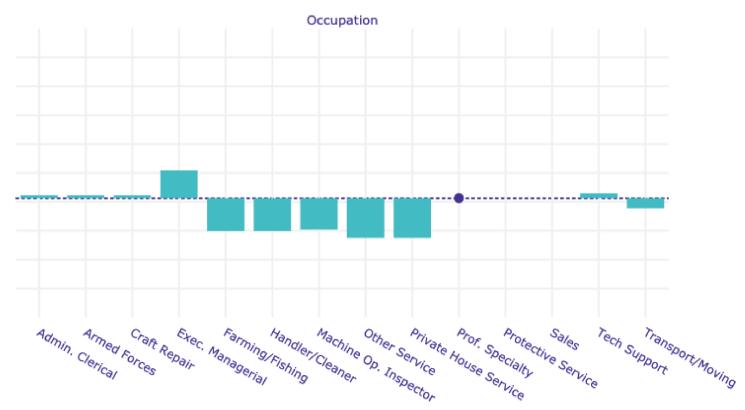
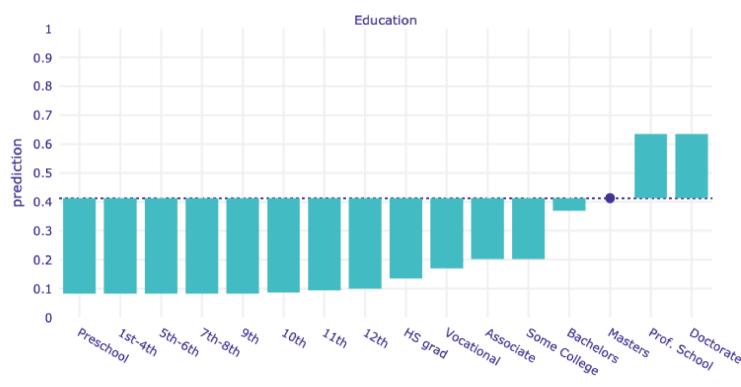
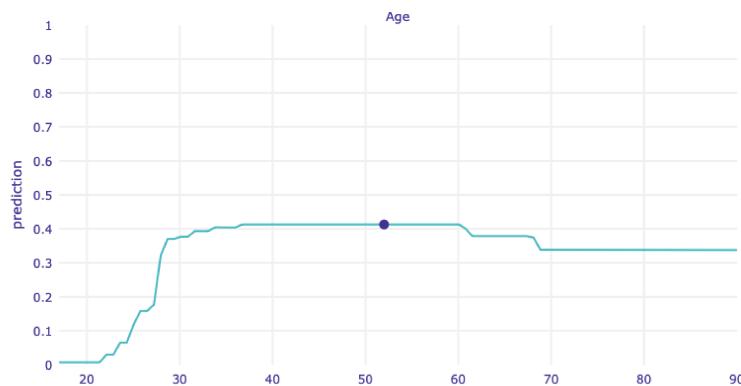
Model Prediction: 0.413



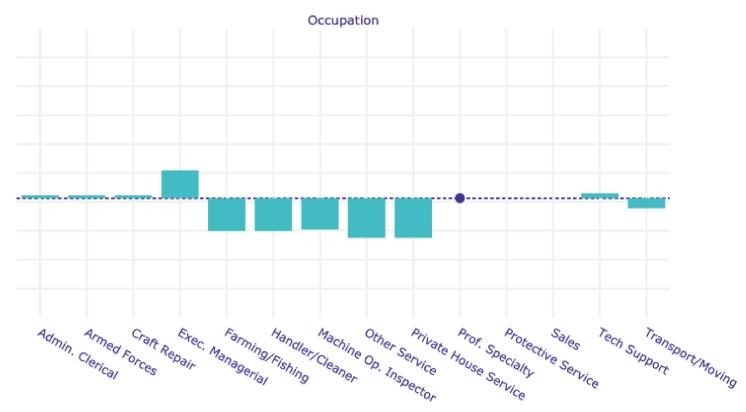
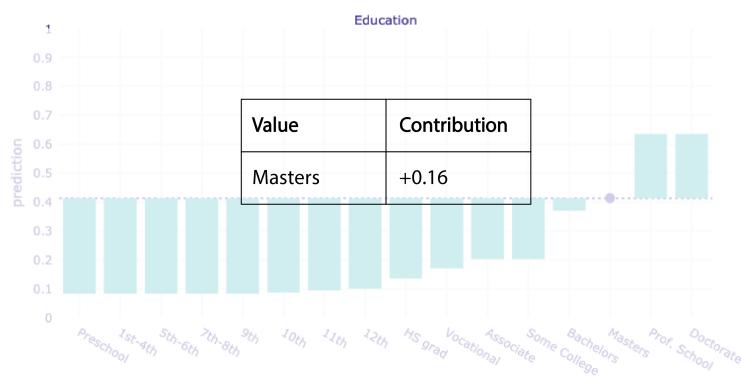
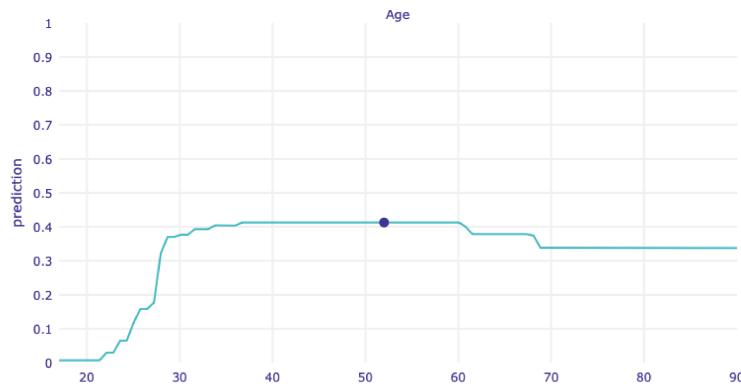
Model Prediction: 0.413



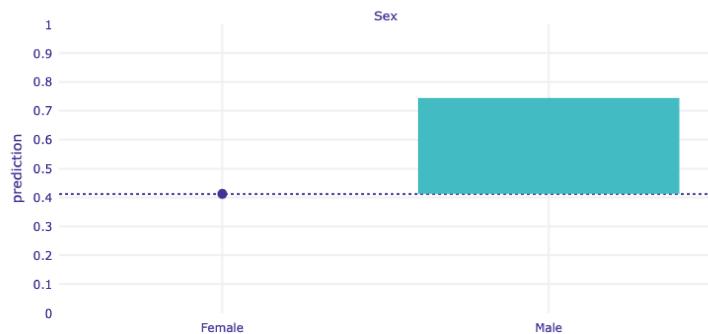
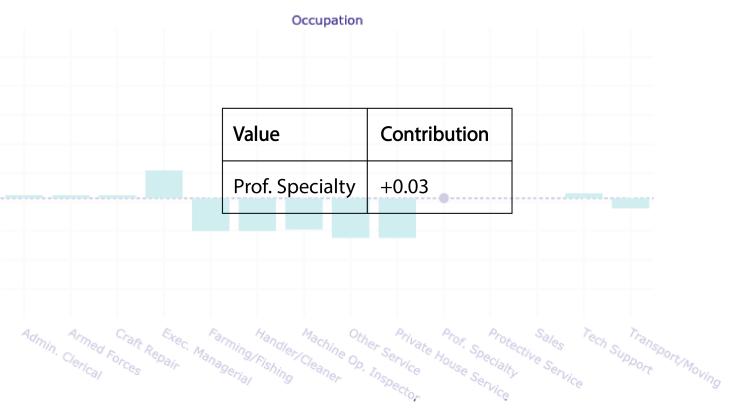
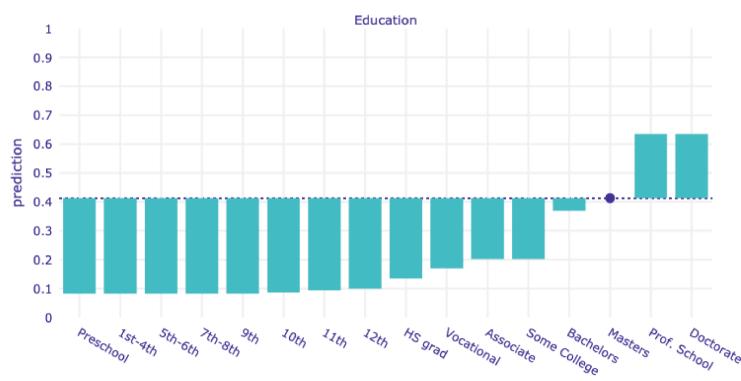
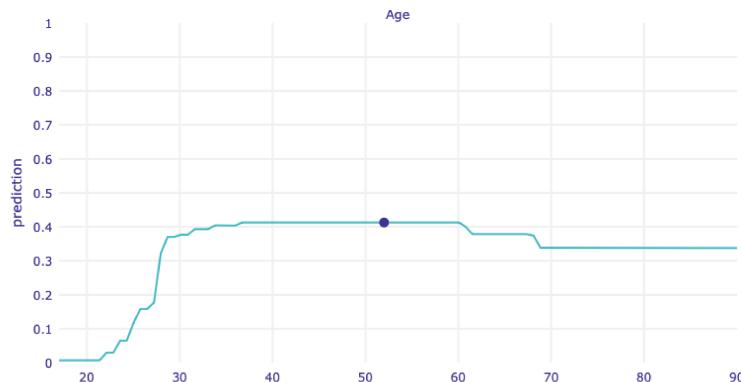
Model Prediction: 0.413



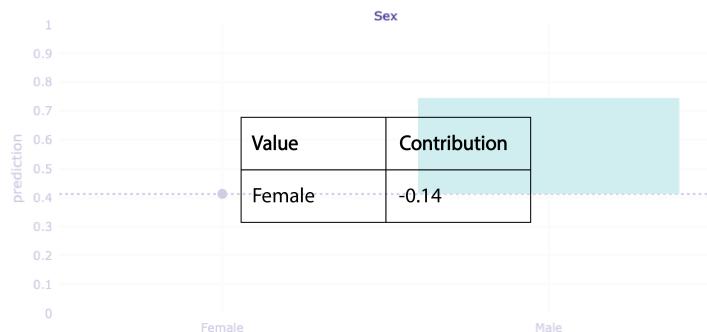
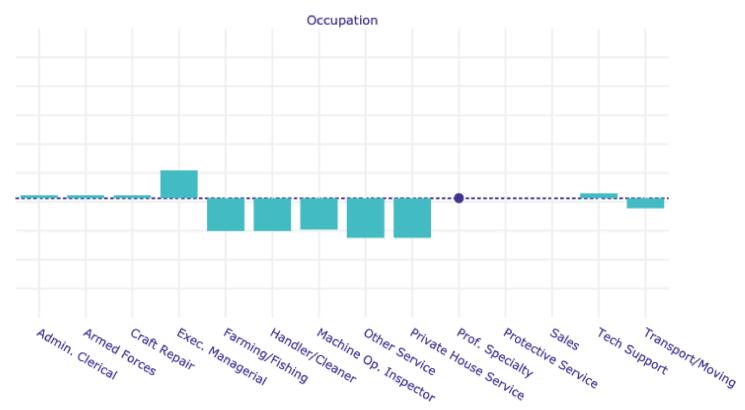
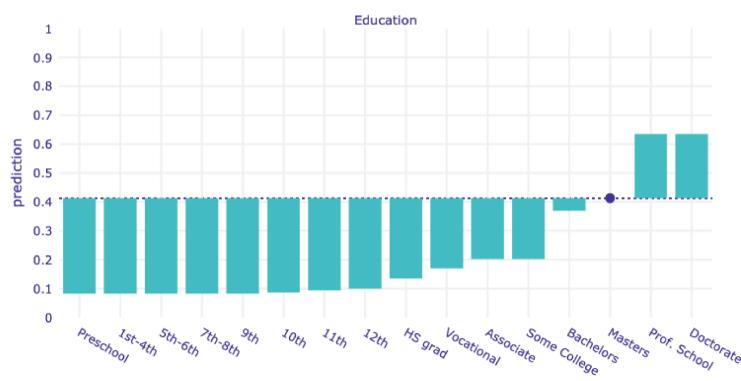
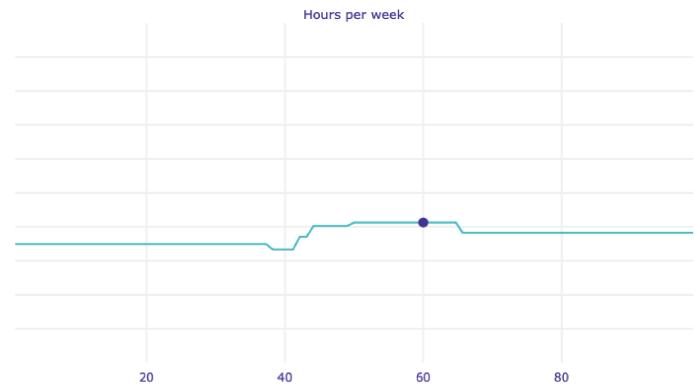
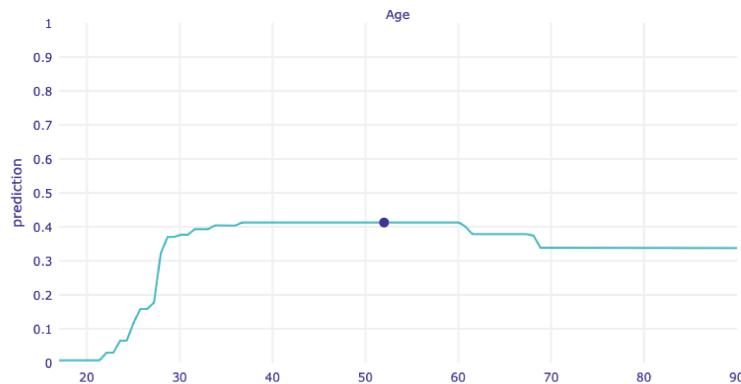
Model Prediction: 0.413



Model Prediction: 0.413

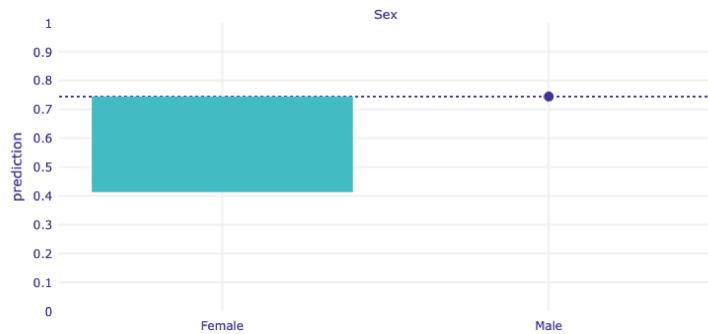
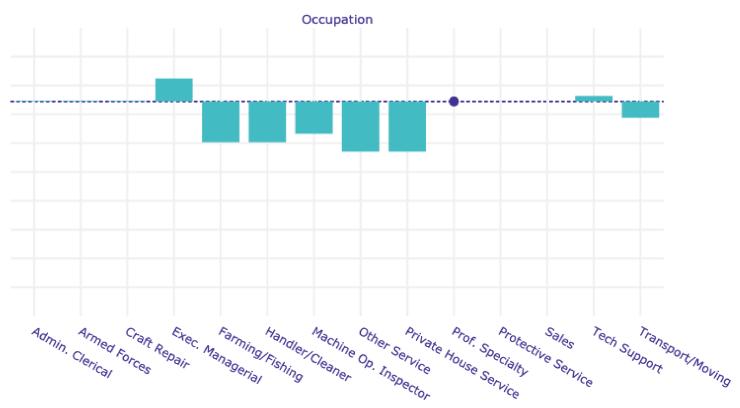
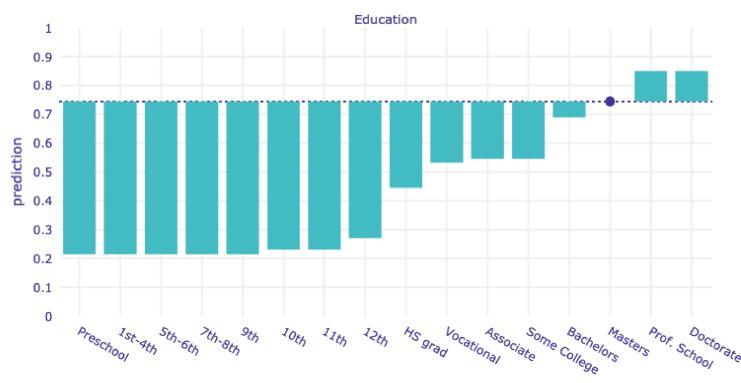
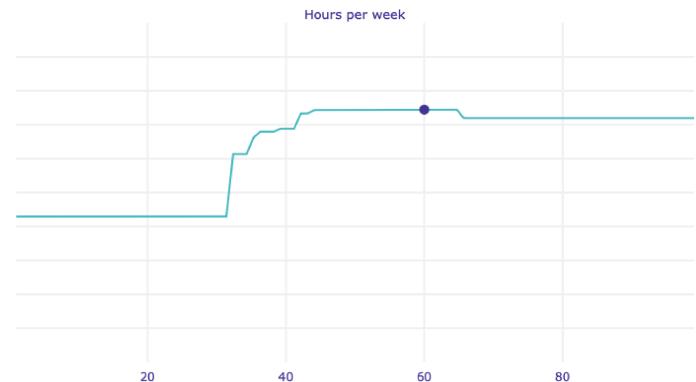
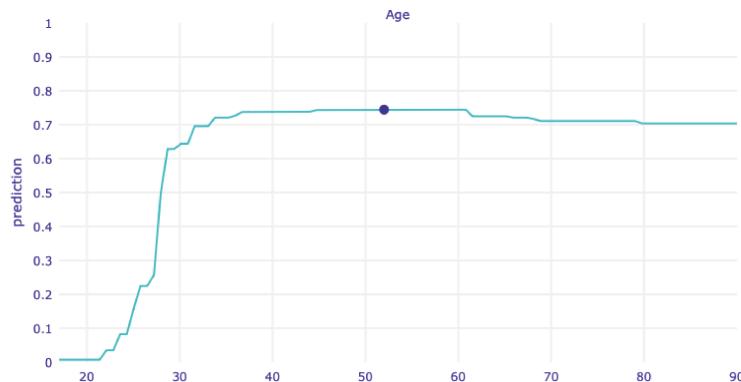


Model Prediction: 0.413

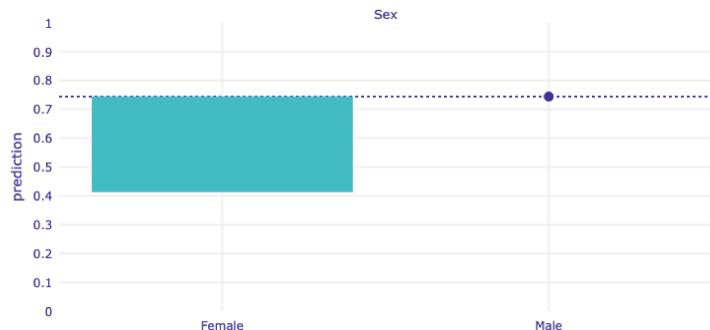
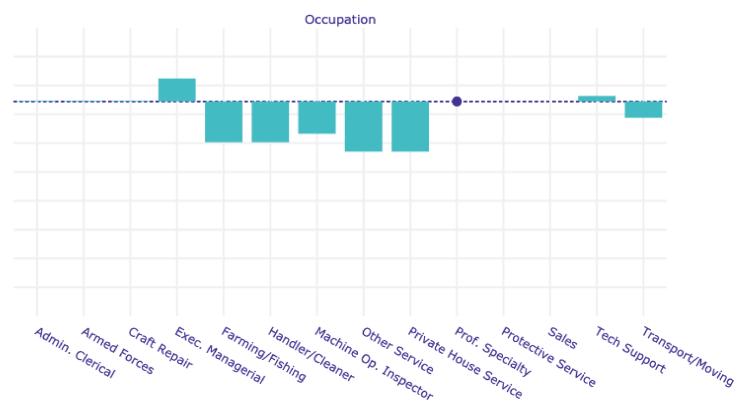
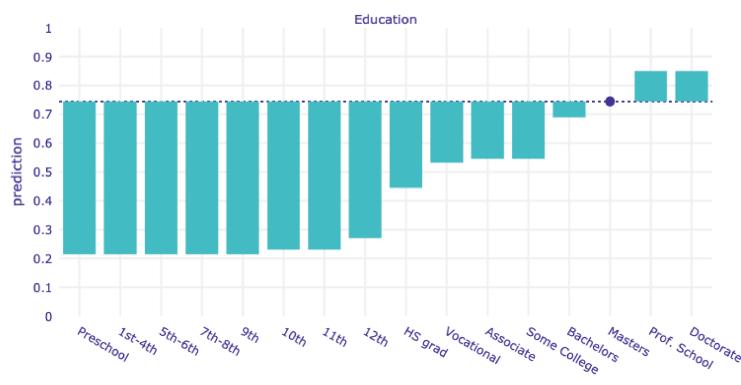
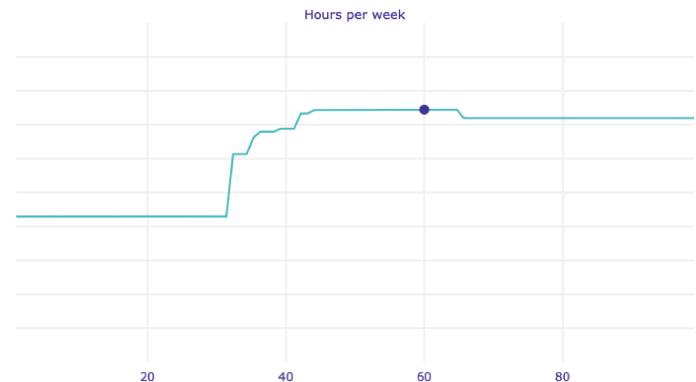
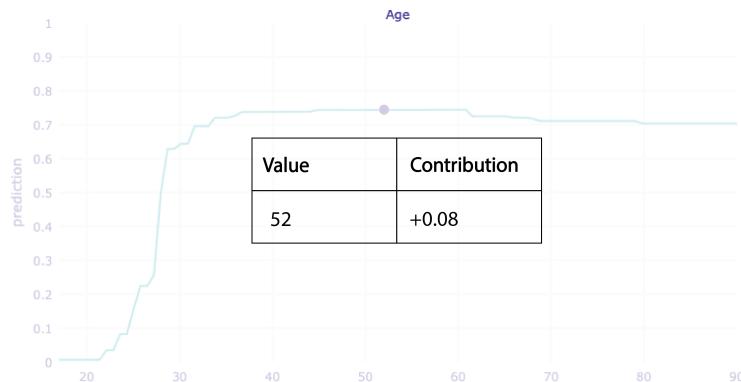


Person B

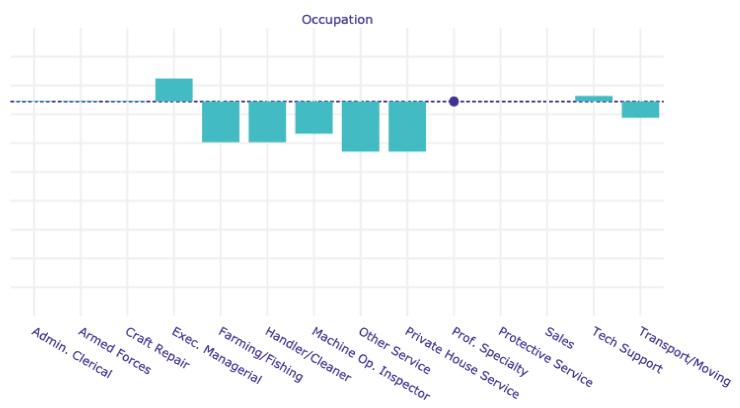
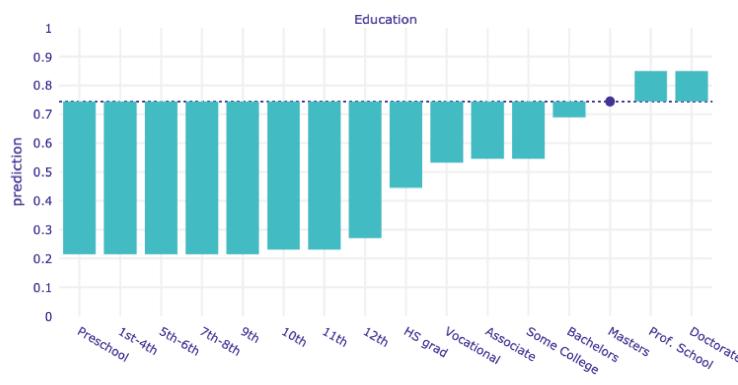
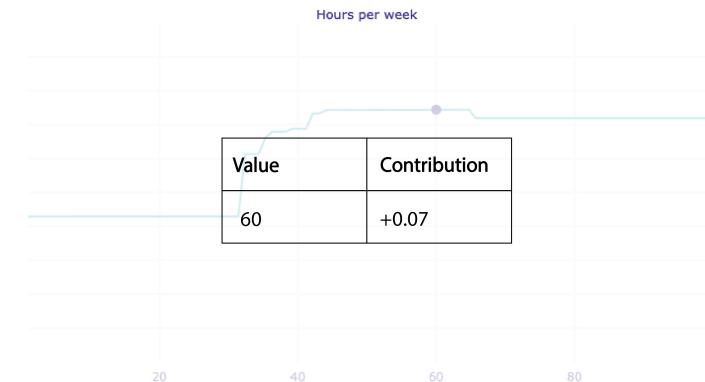
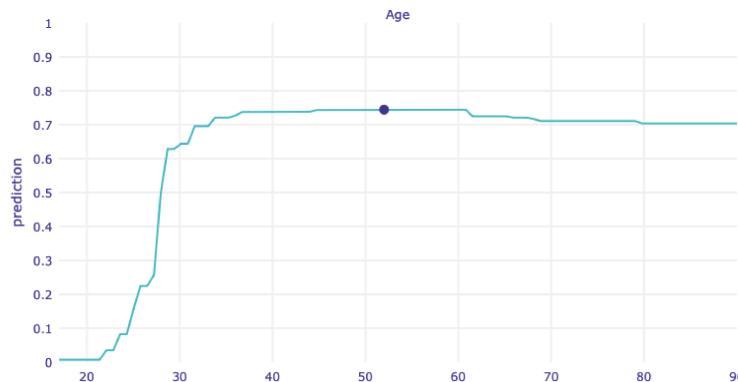
Model Prediction: 0.745



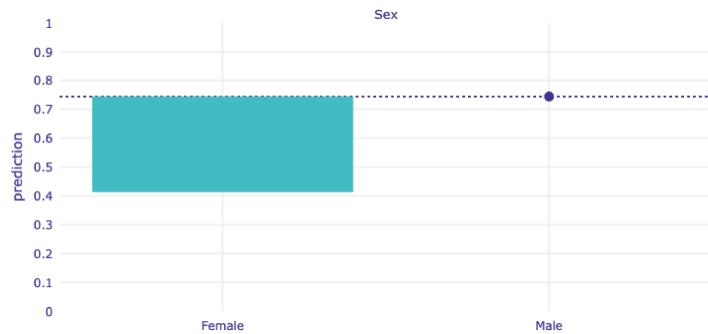
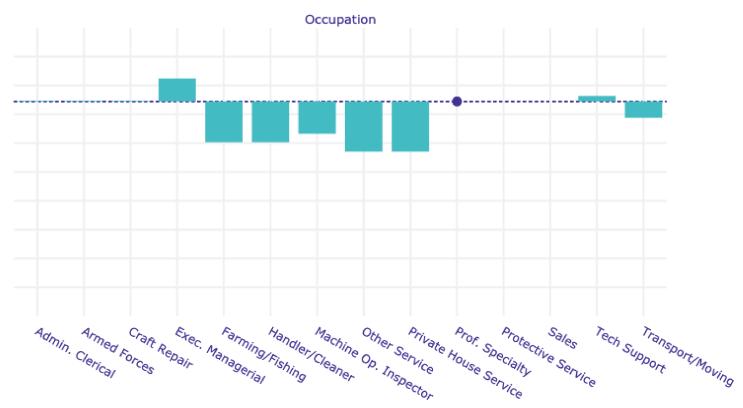
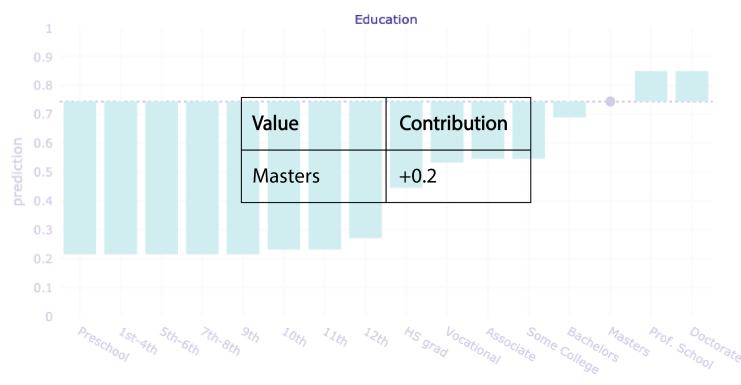
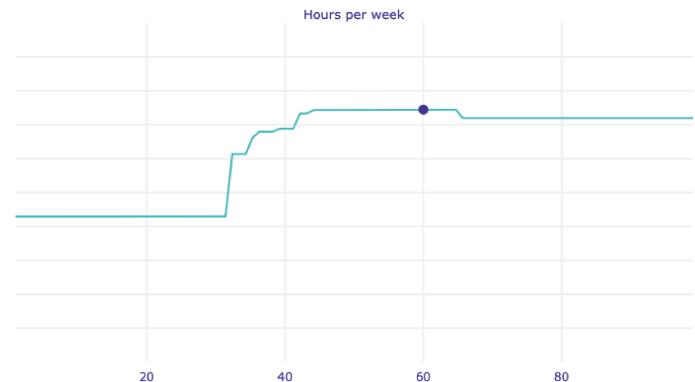
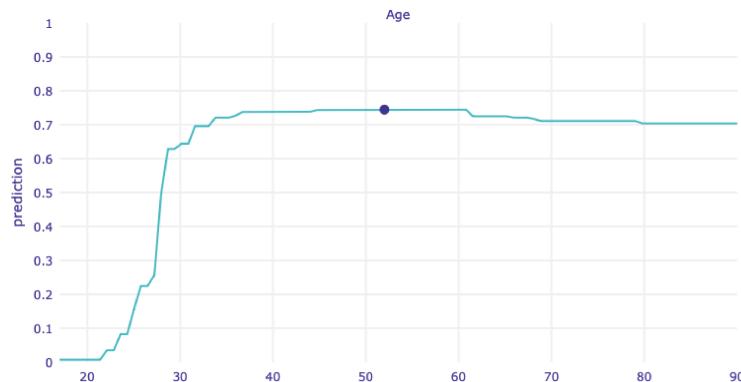
Model Prediction: 0.745



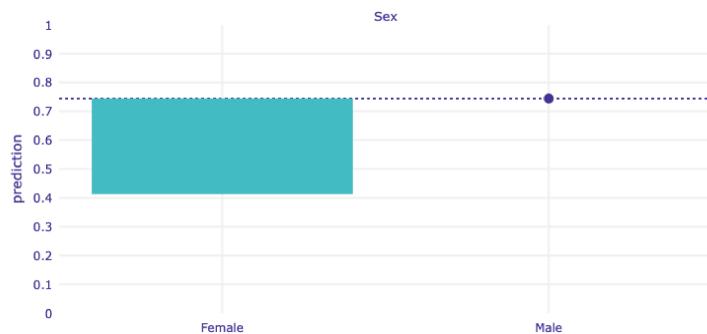
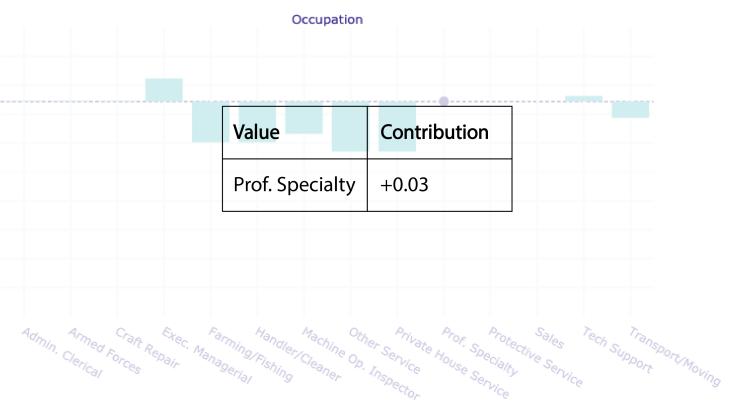
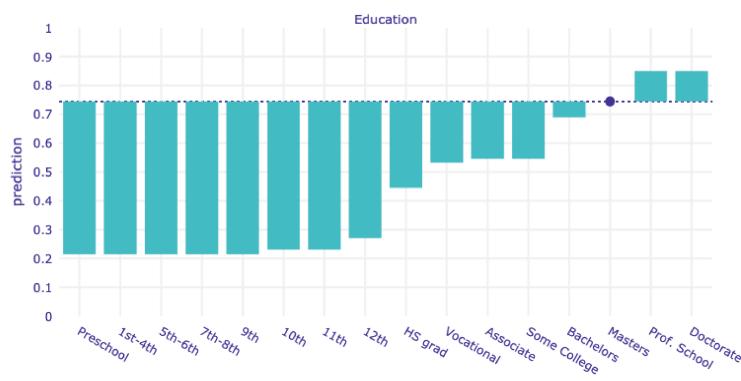
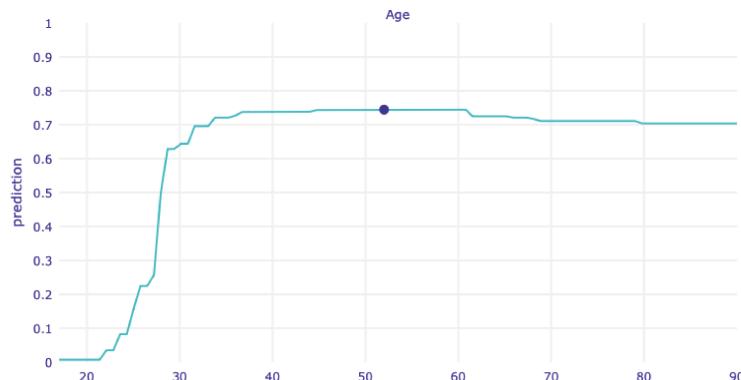
Model Prediction: 0.745



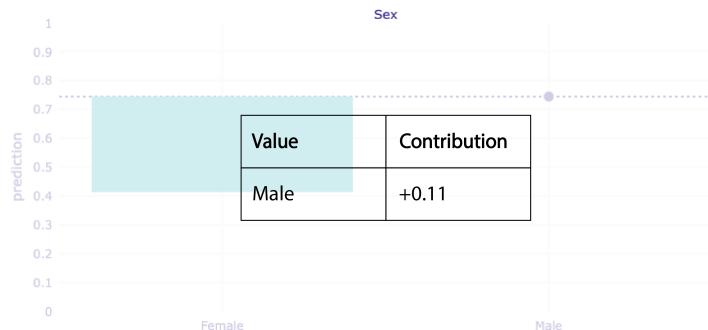
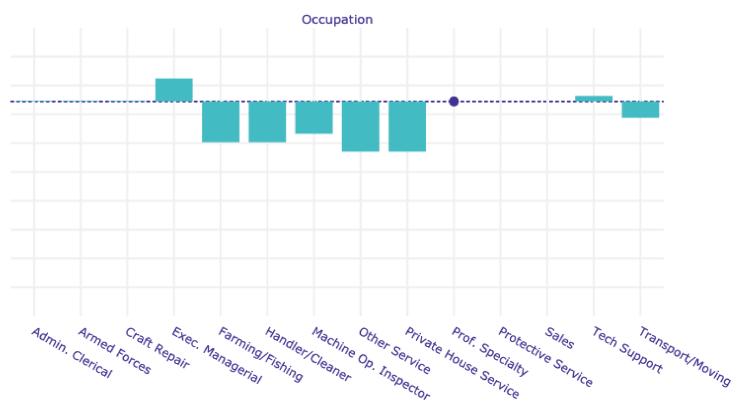
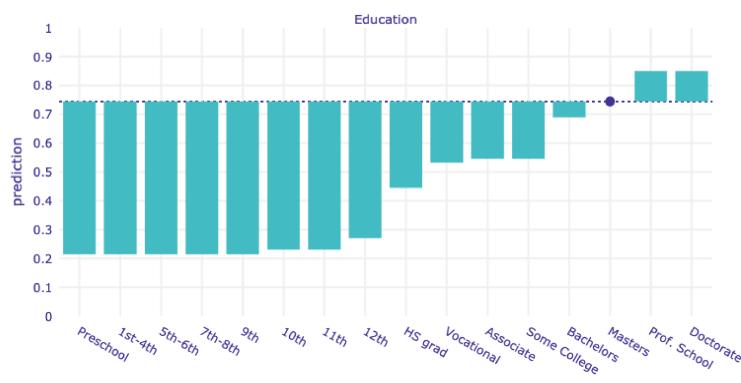
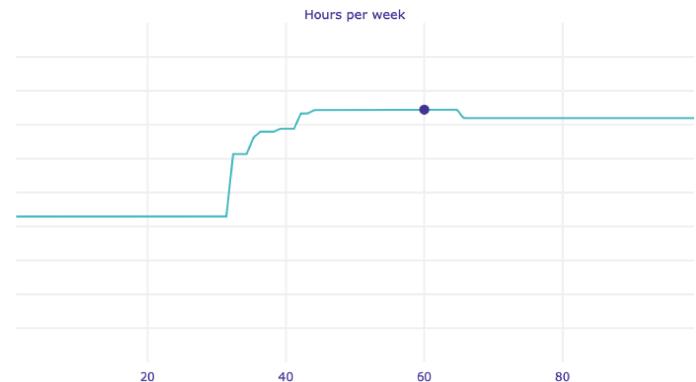
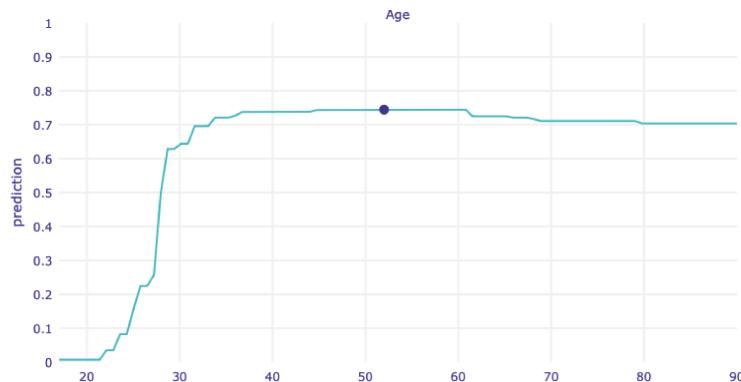
Model Prediction: 0.745



Model Prediction: 0.745



Model Prediction: 0.745



Will this model approve the loan for **Person A**?

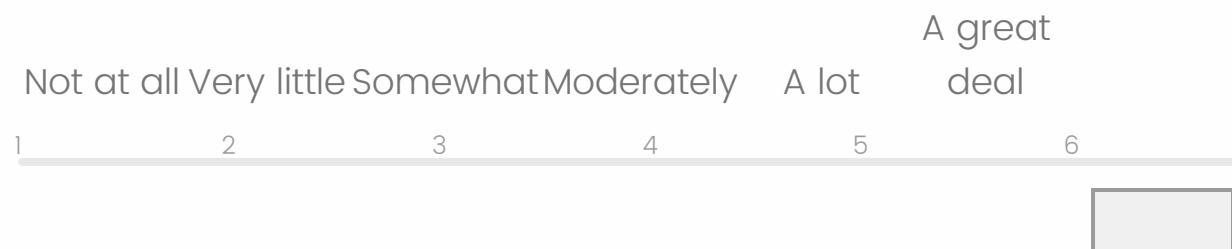
YES

NO

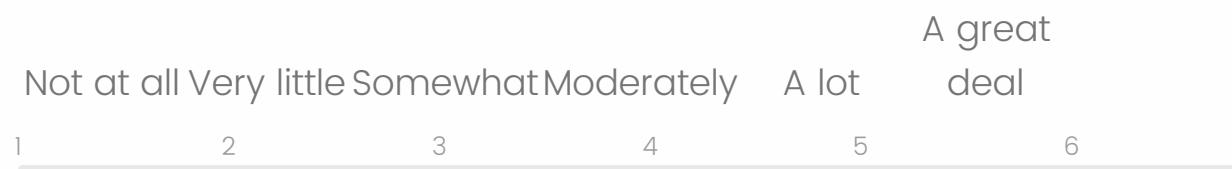
Will this model approve the loan for **Person B**?

 YES NO

On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **you**?



On a scale from 1 to 6, how much do you trust the model to approve or deny a loan for **other people in general**?



Level of Trust



Please indicate whether you agree with the below statements.

Agree

This model uses all of the features that it should use when making this decision.

This model does not use any unnecessary features when making this decision.

I trust the data this model was trained on.

Computer models can be trusted to make human decisions.

This model is accurate.

This model is fair.

This model would probably give me a loan because I am similar to a person described in this question.

This model would probably give me a loan because I am different from a person described in this question.

This model would probably give me a loan because of previous decisions it has made.

This model probably would not give me a loan, and this would be the correct decision.

Fairness General

Person A

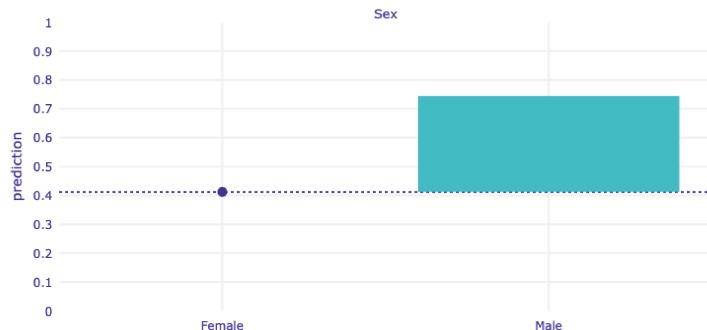
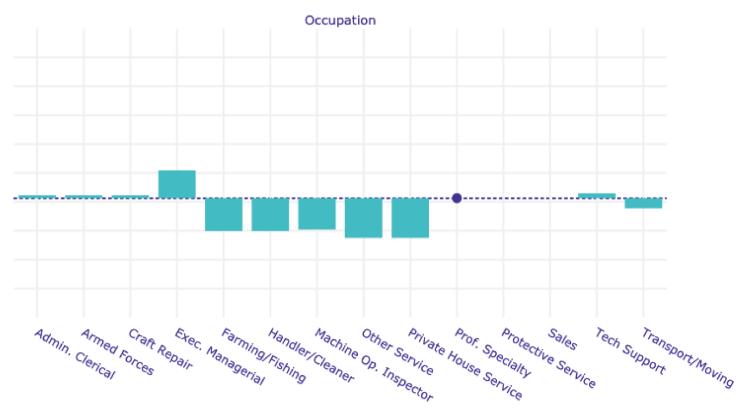
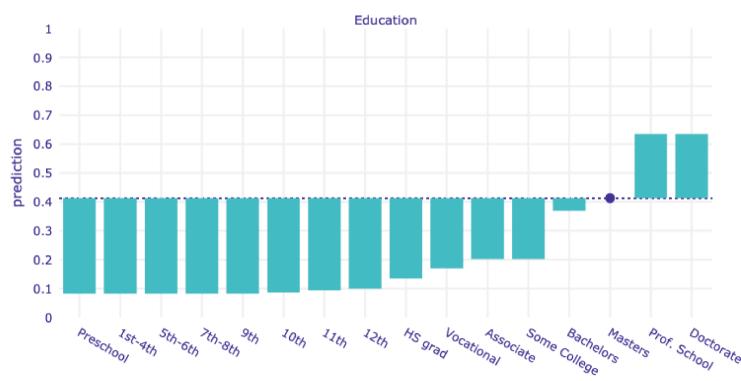
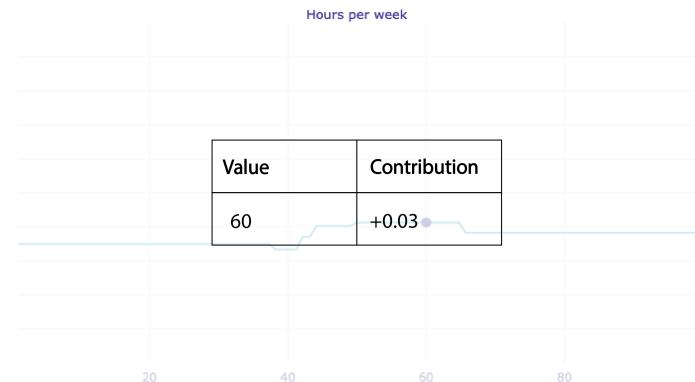
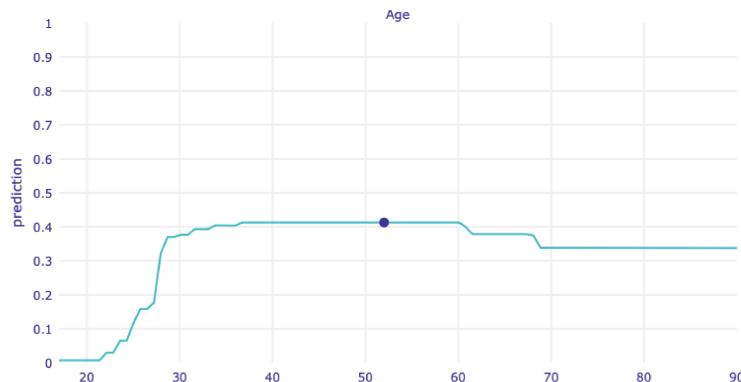
Model Prediction: 0.413



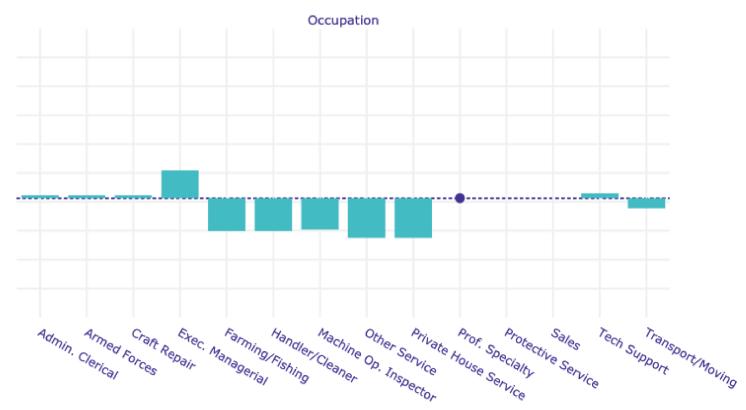
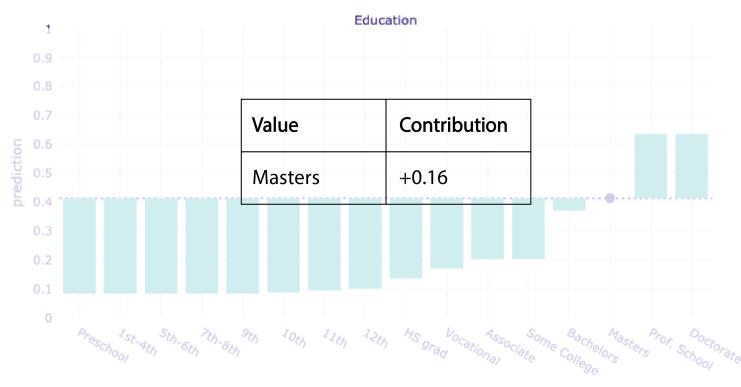
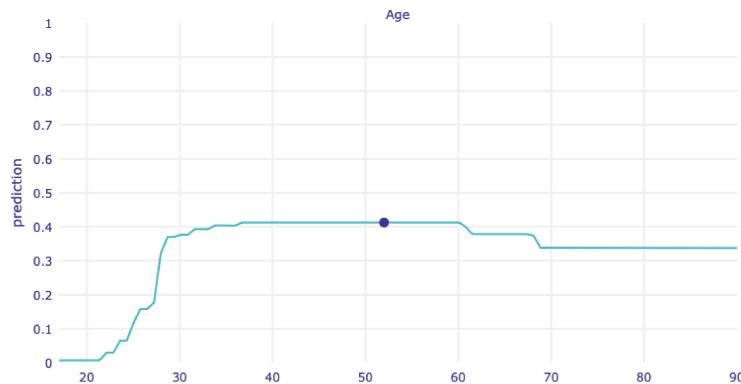
Model Prediction: 0.413



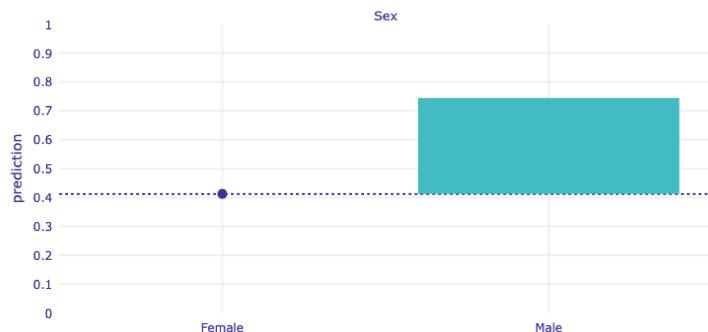
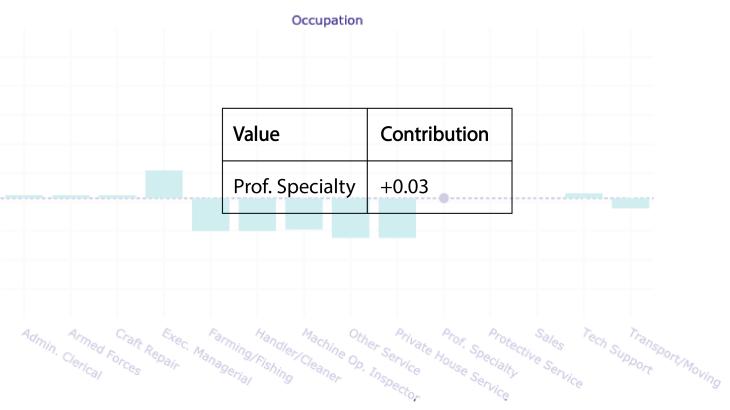
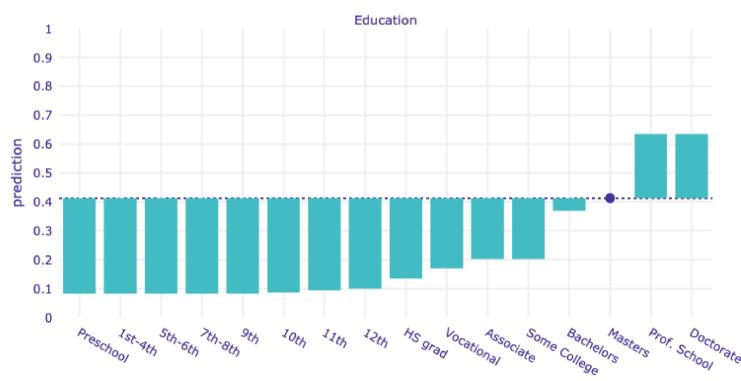
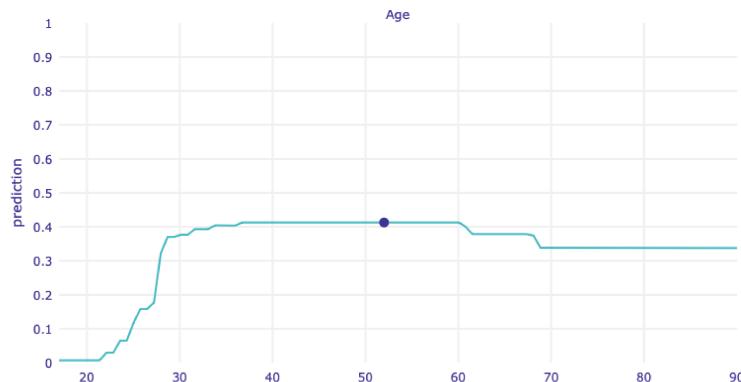
Model Prediction: 0.413



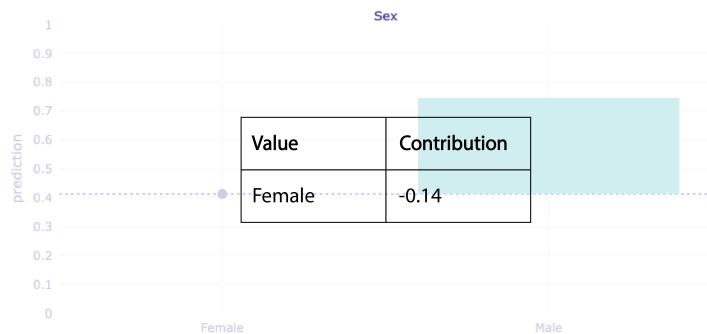
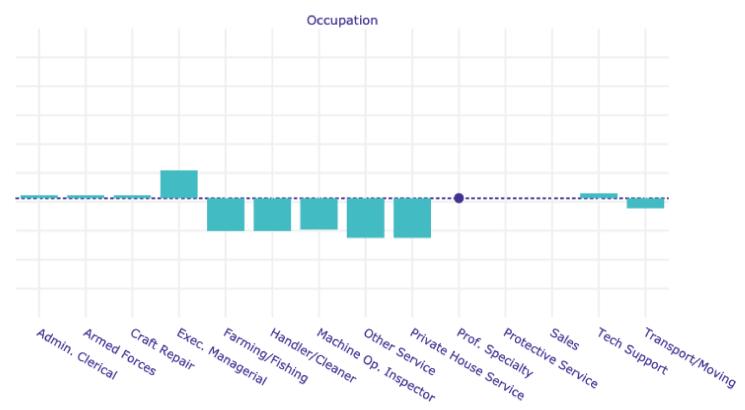
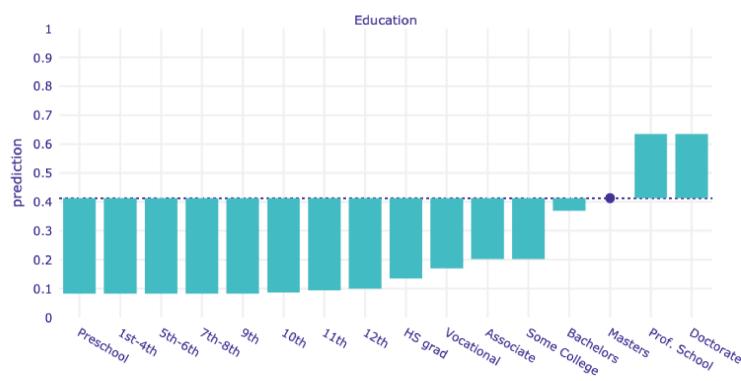
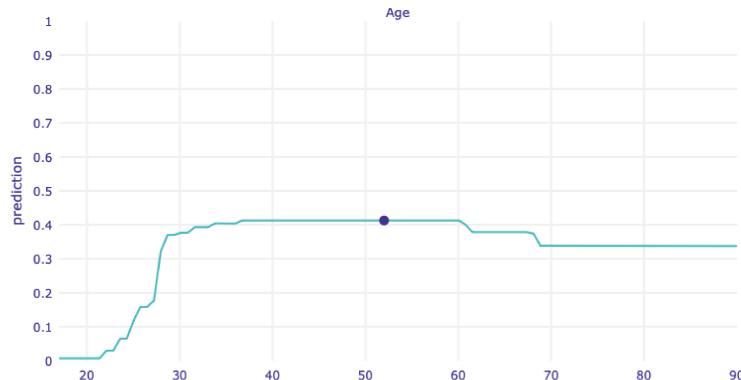
Model Prediction: 0.413



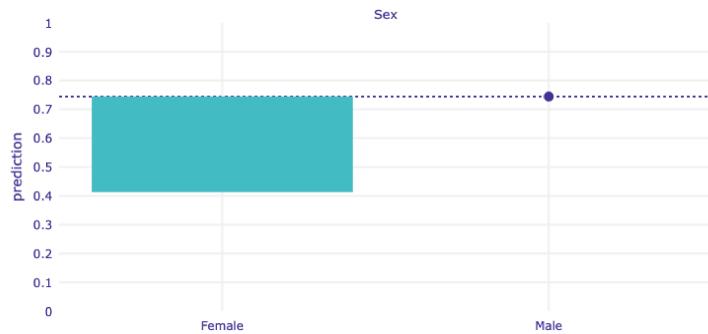
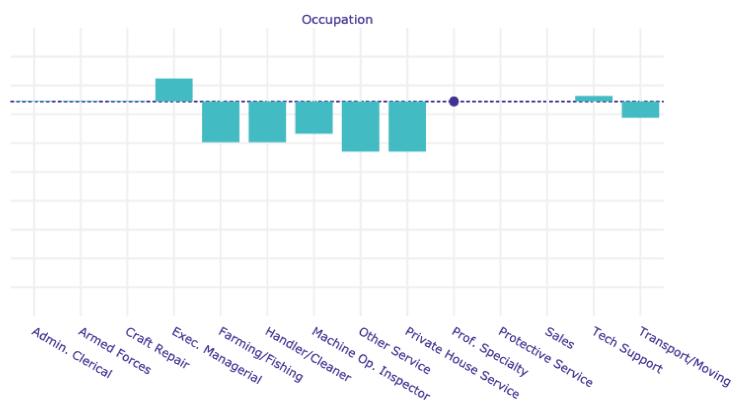
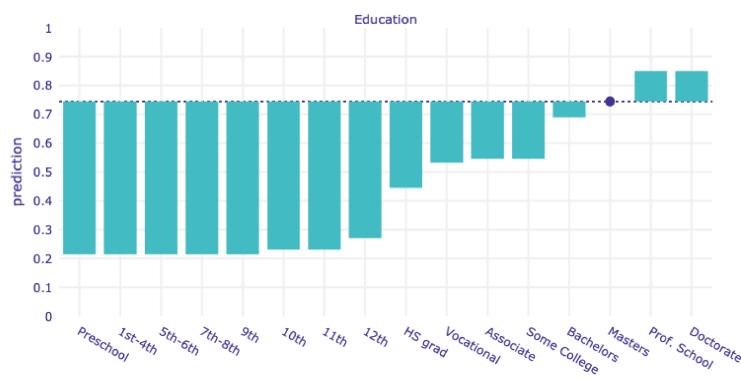
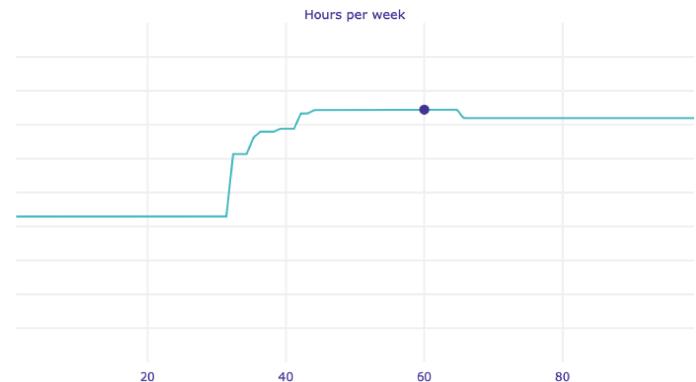
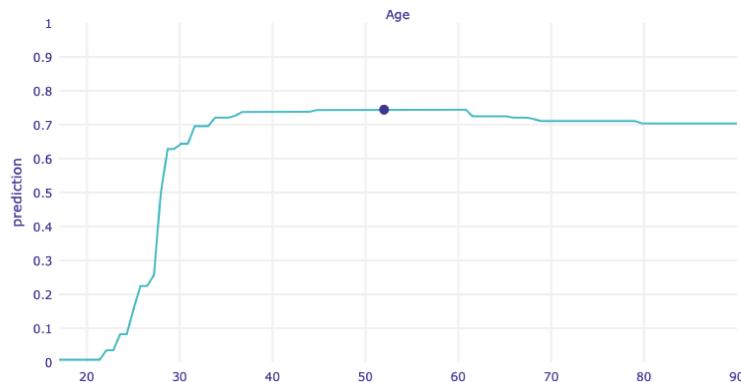
Model Prediction: 0.413



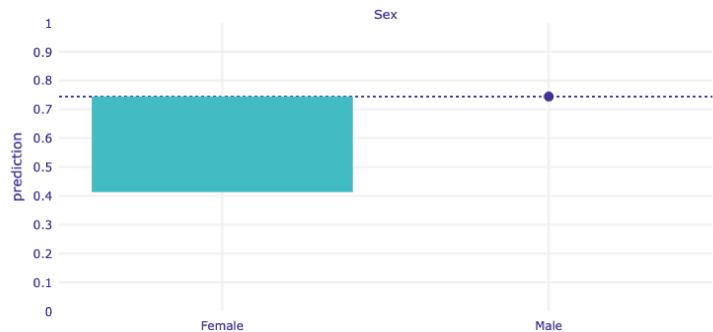
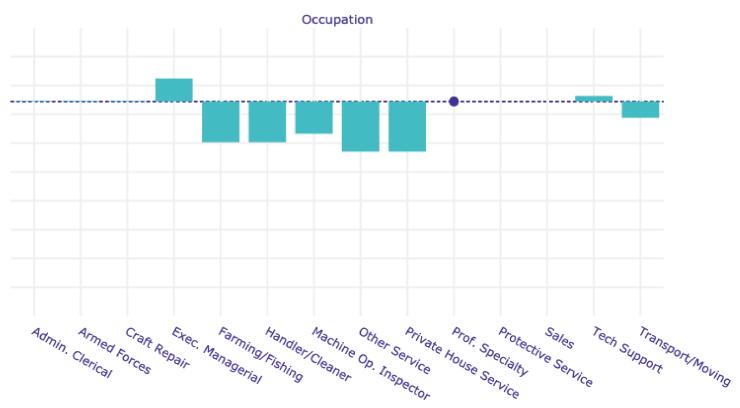
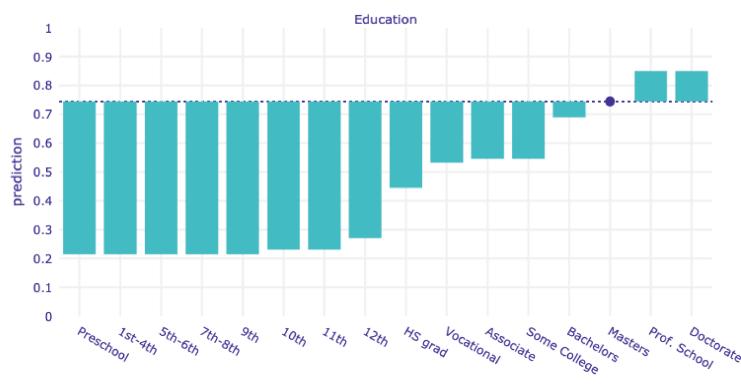
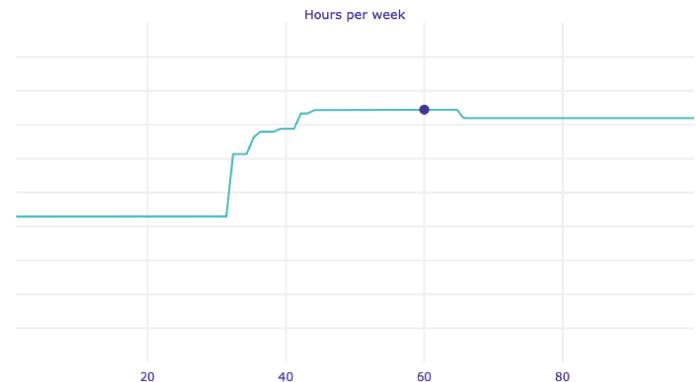
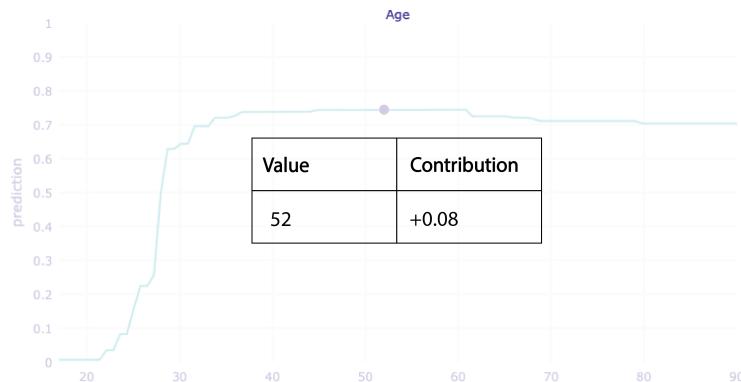
Model Prediction: 0.413



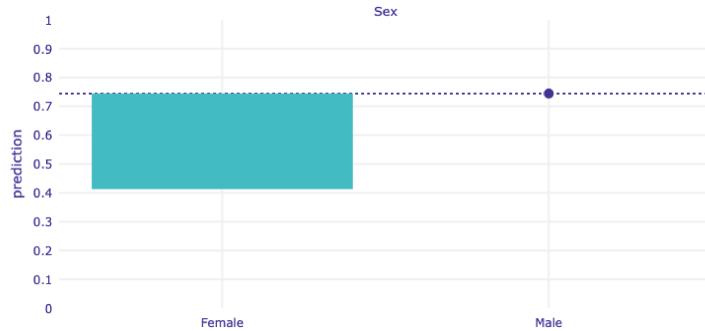
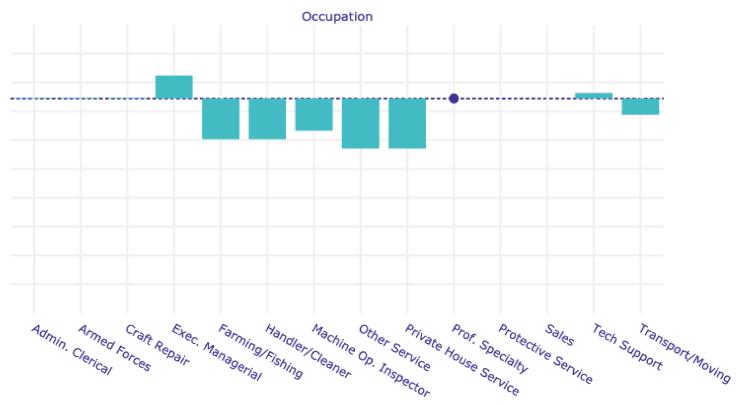
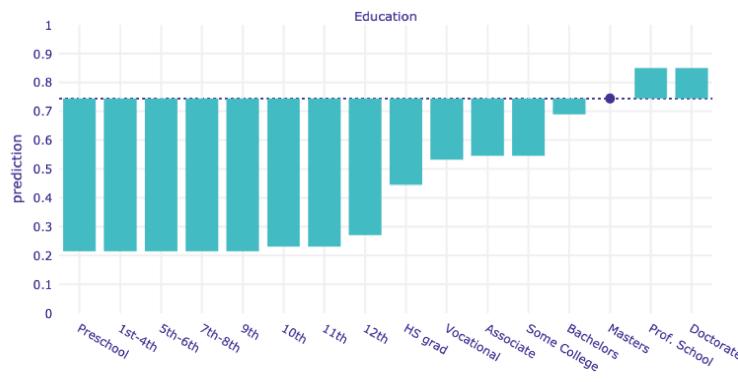
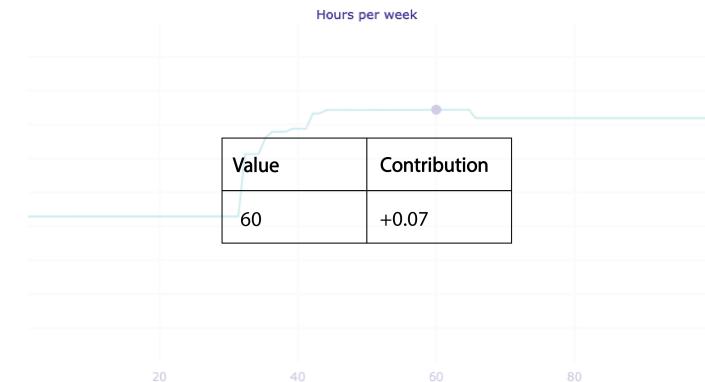
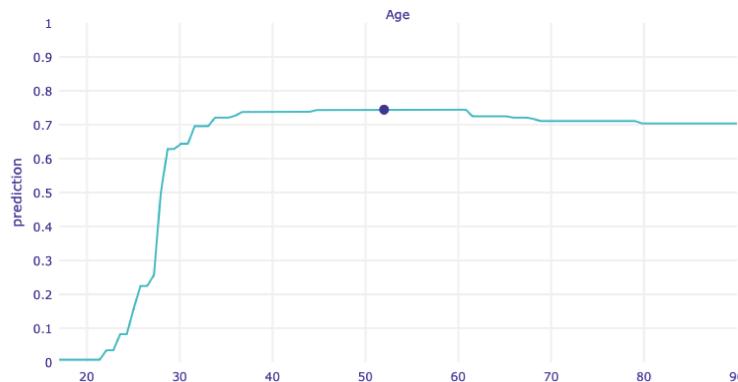
Person B

Model Prediction: 0.745

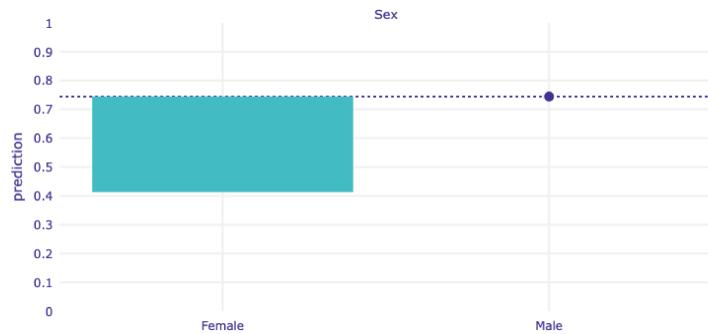
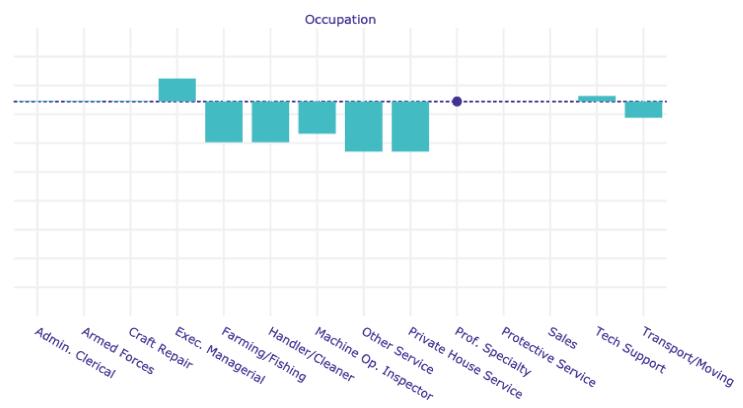
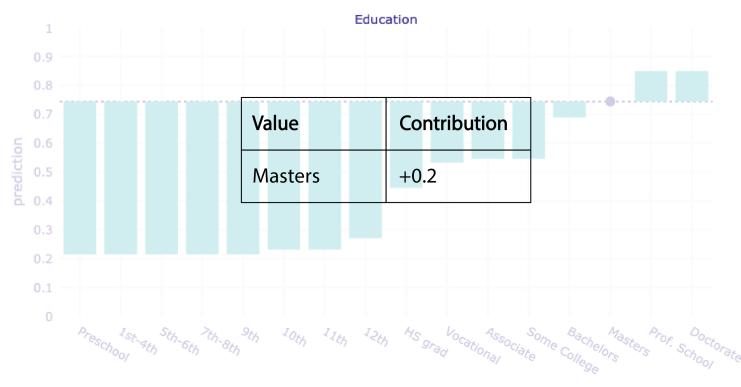
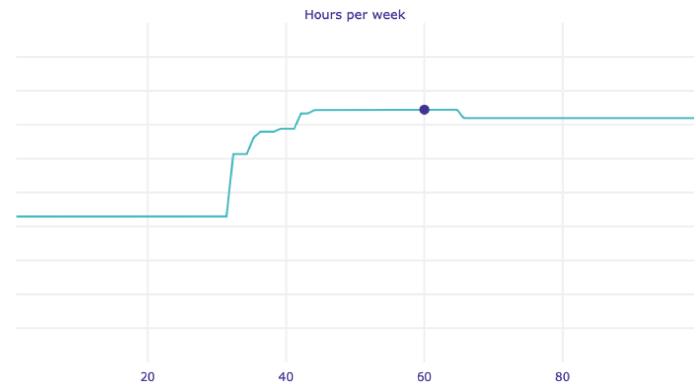
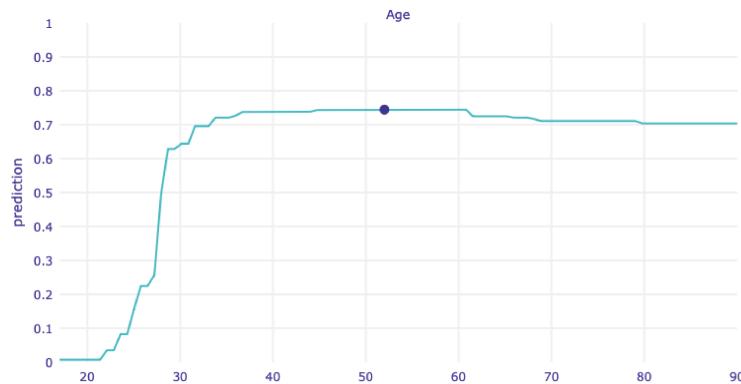
Model Prediction: 0.745



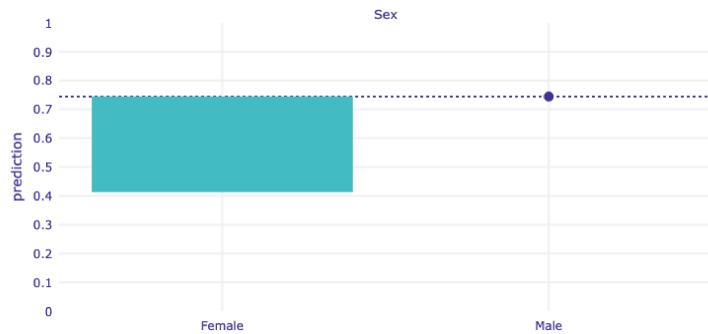
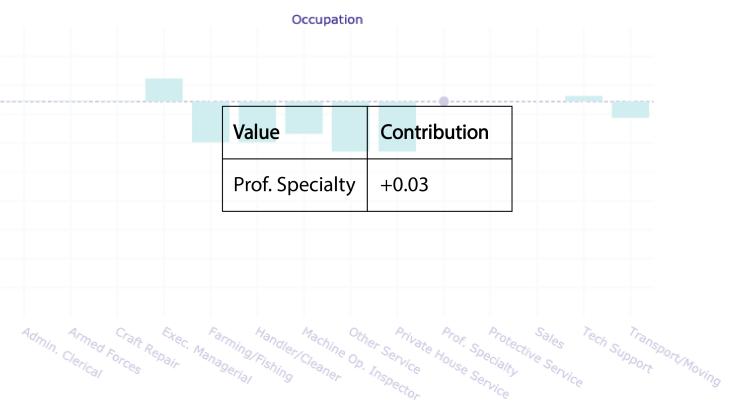
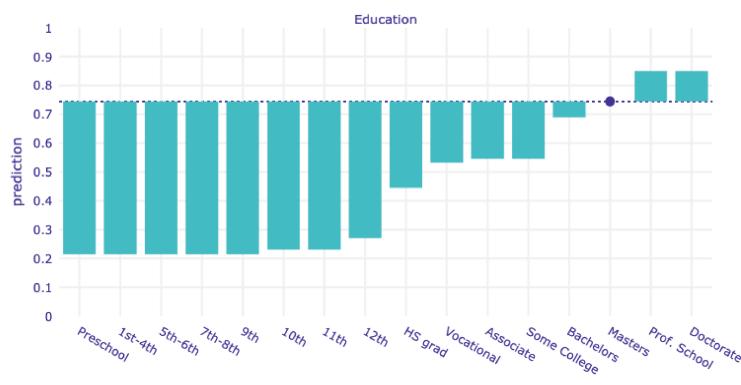
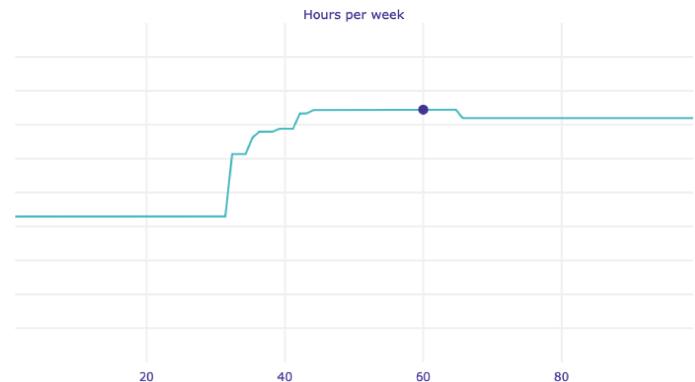
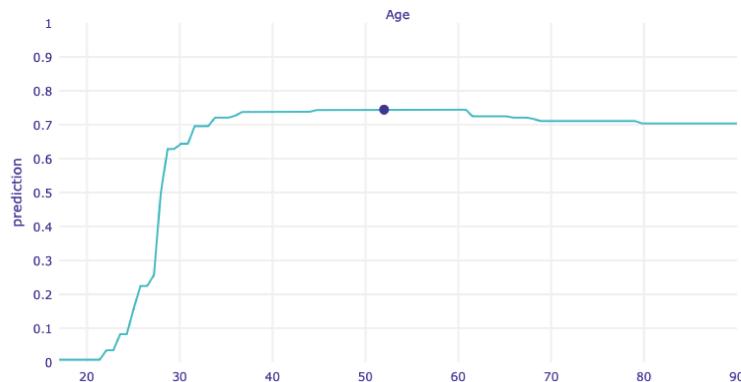
Model Prediction: 0.745



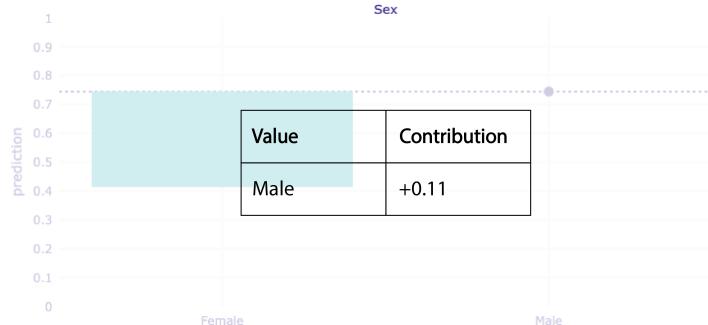
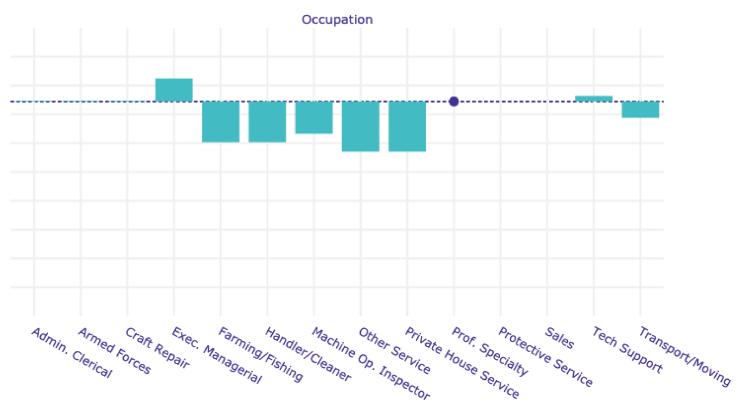
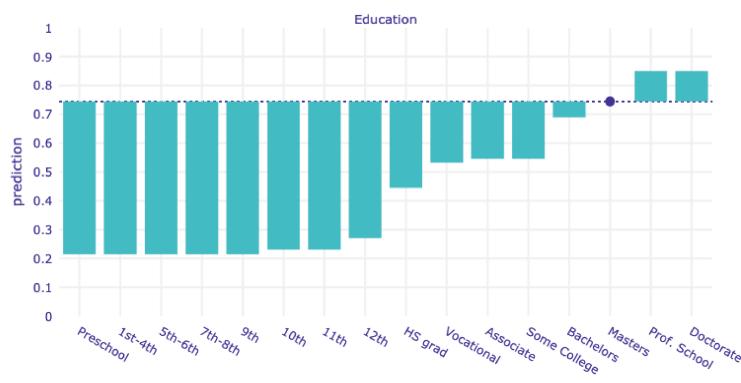
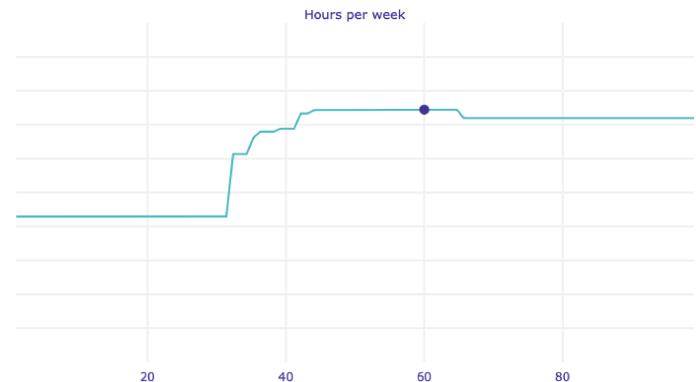
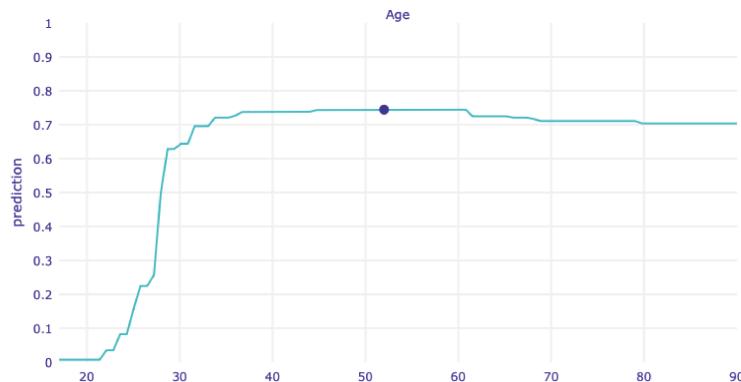
Model Prediction: 0.745



Model Prediction: 0.745



Model Prediction: 0.745



Do you think this model includes potentially discriminating factors?

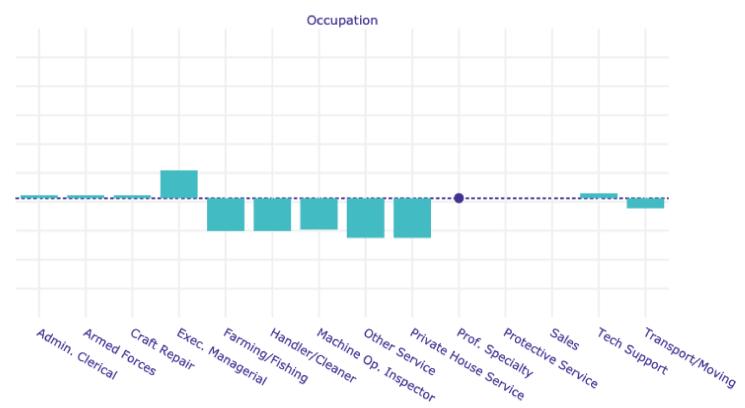
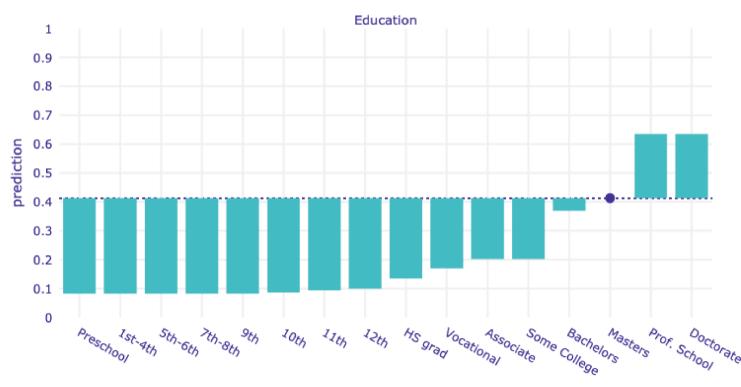
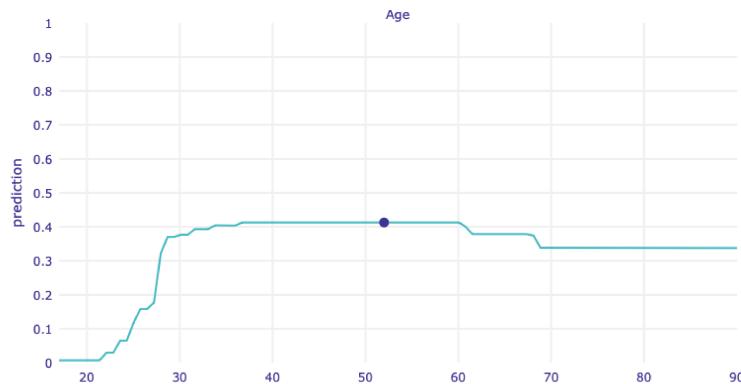
YES NO

If yes, which ones?

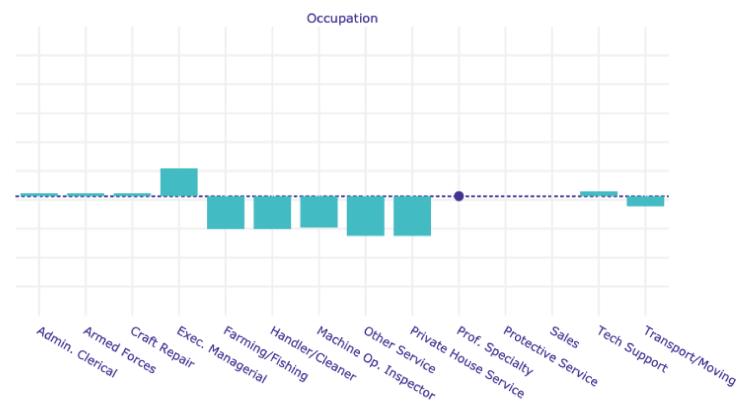
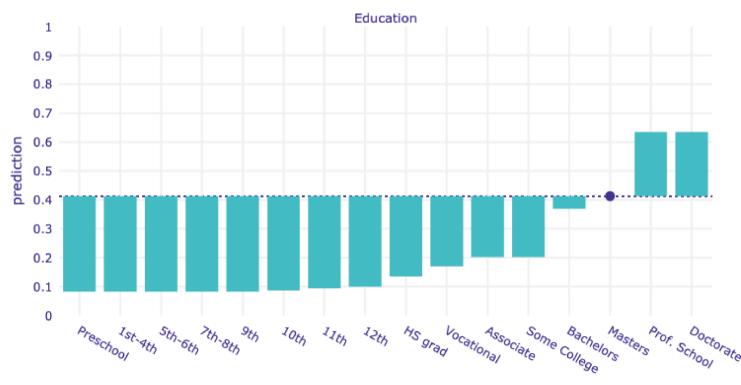
- Age
- Hours Per Week
- Education
- Occupation
- Sex

Person A

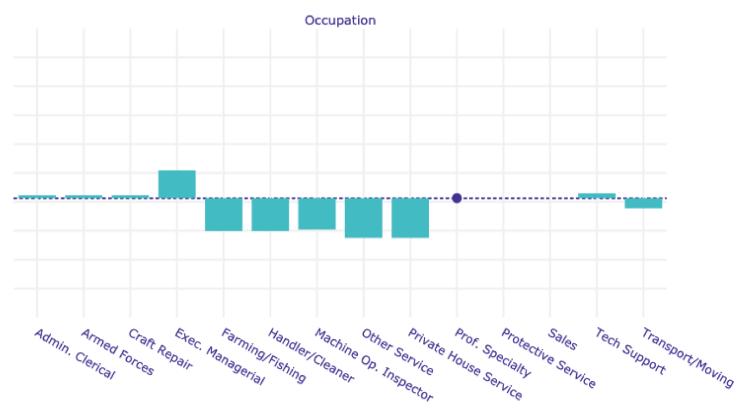
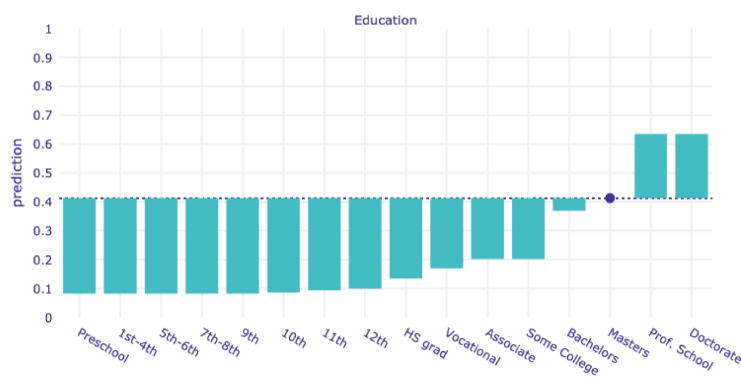
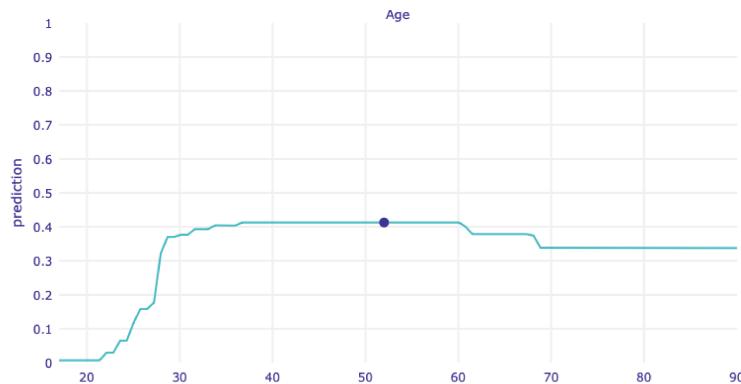
Model Prediction: 0.413



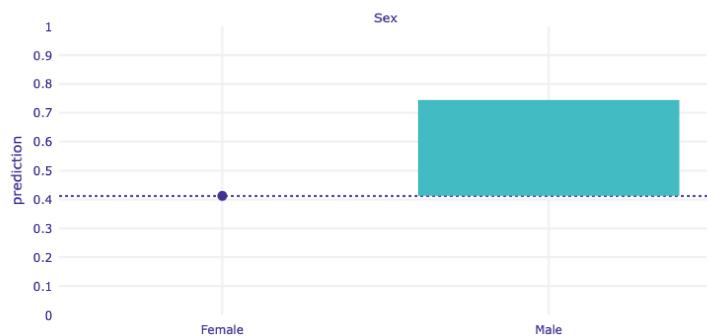
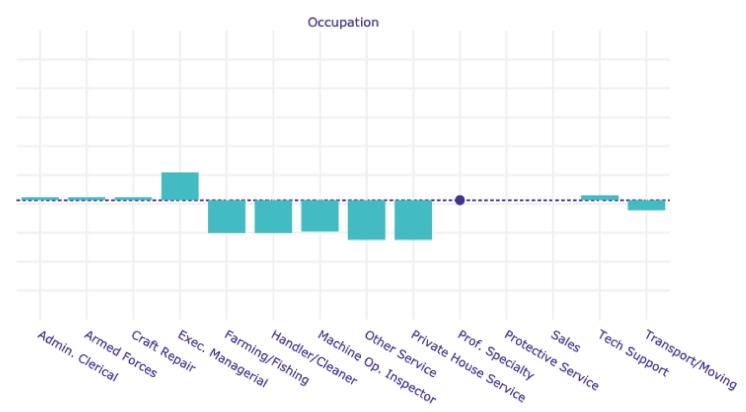
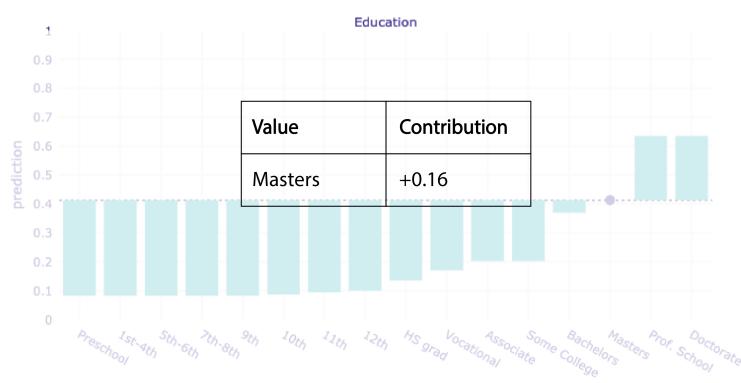
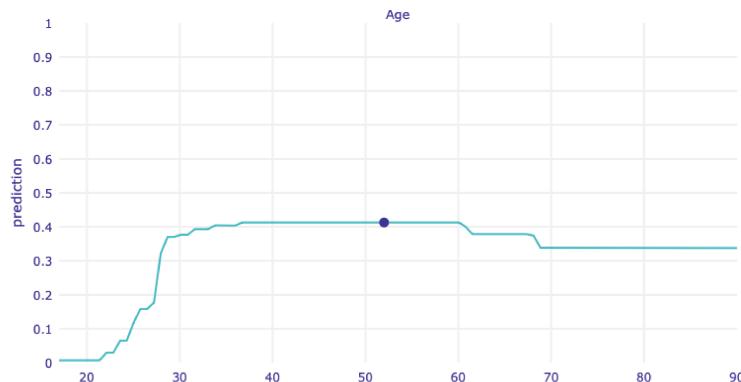
Model Prediction: 0.413



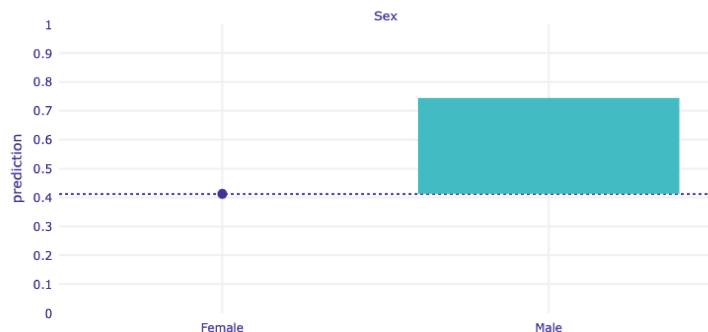
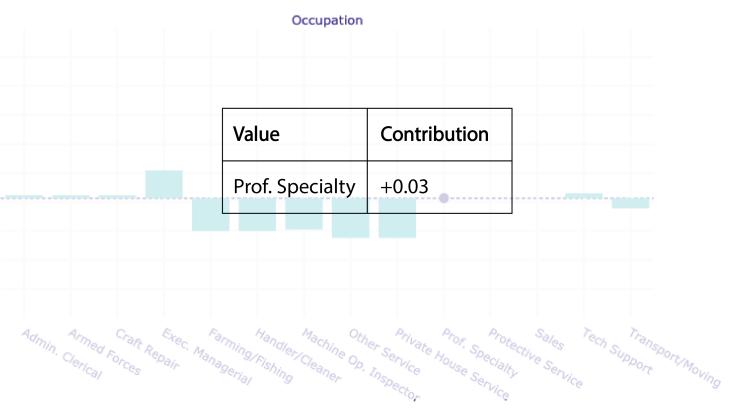
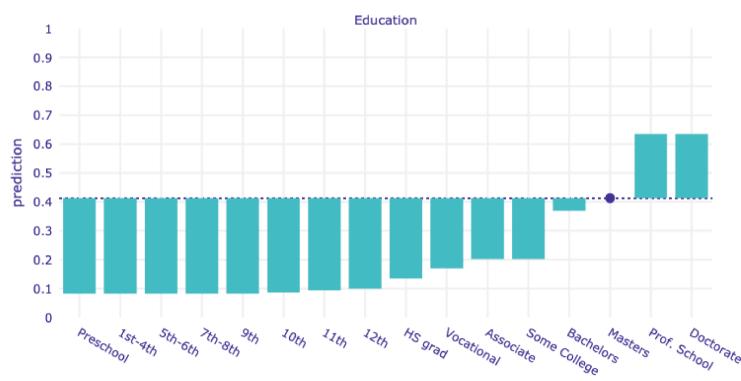
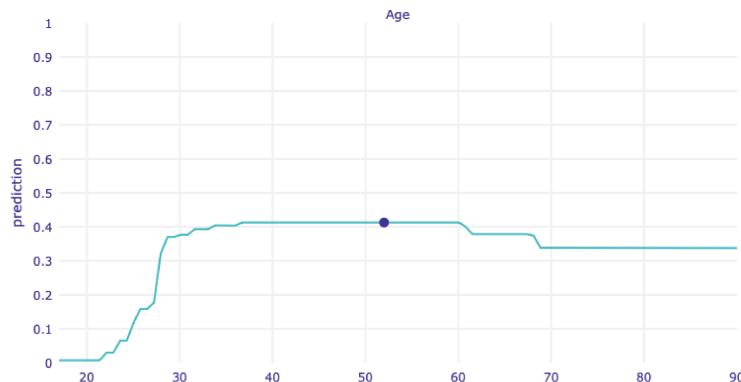
Model Prediction: 0.413



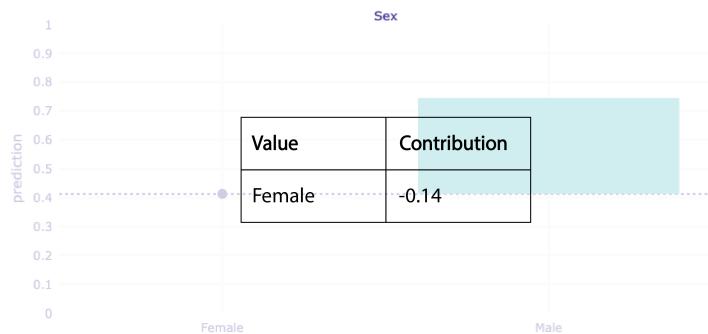
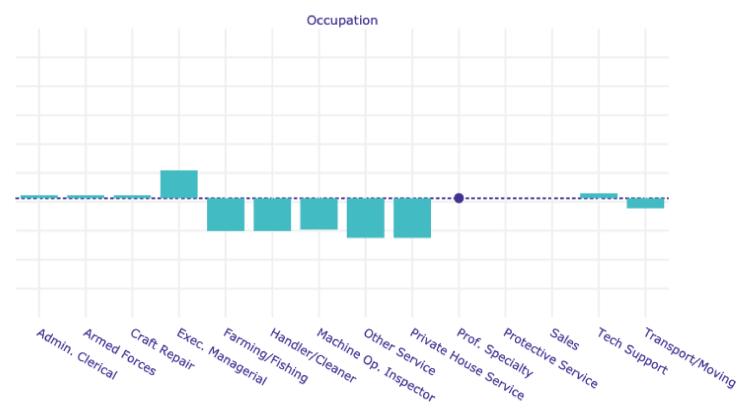
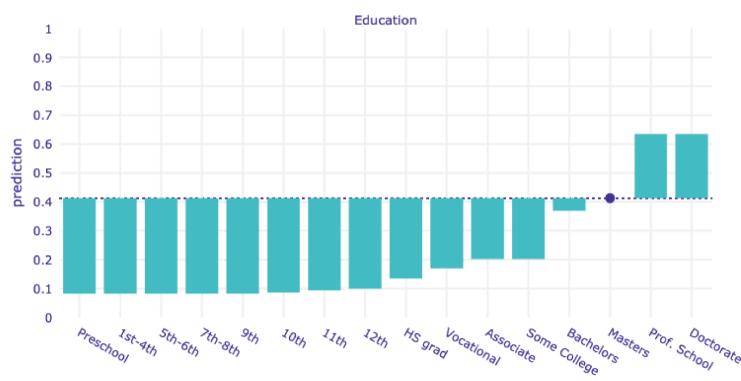
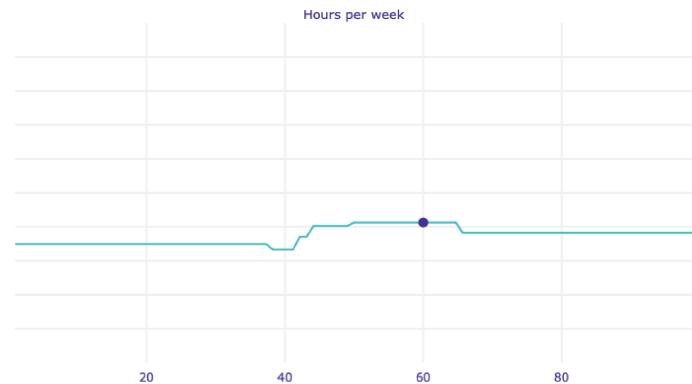
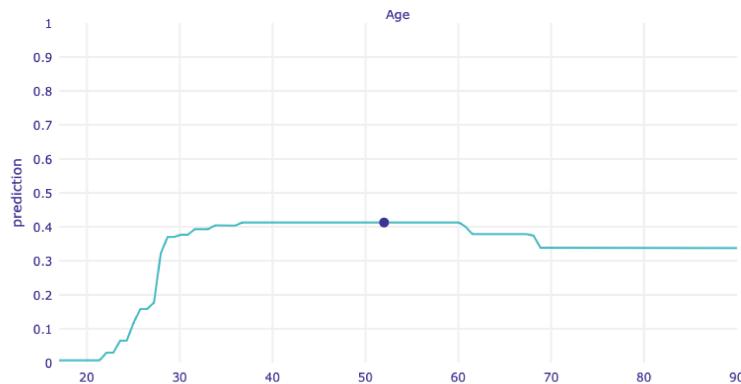
Model Prediction: 0.413



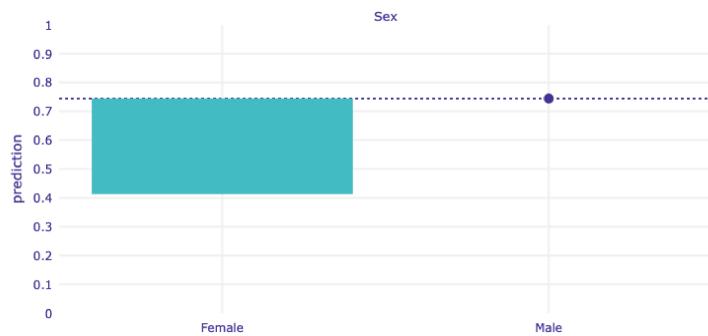
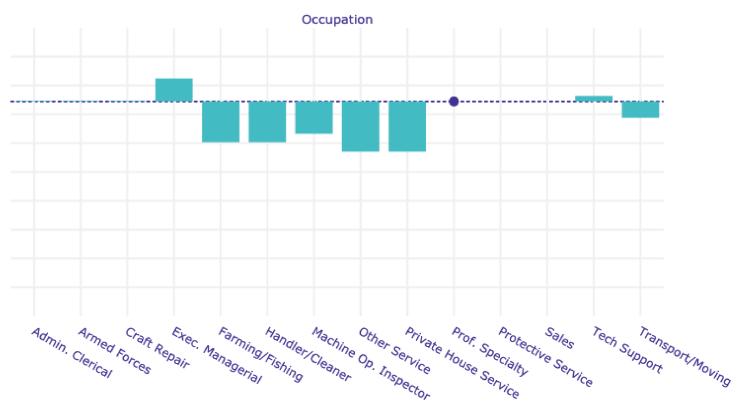
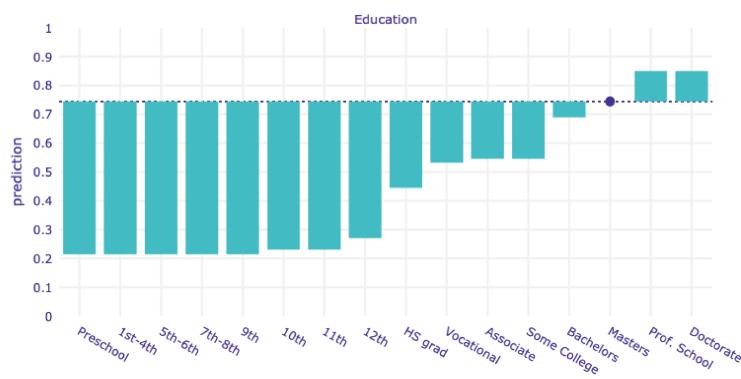
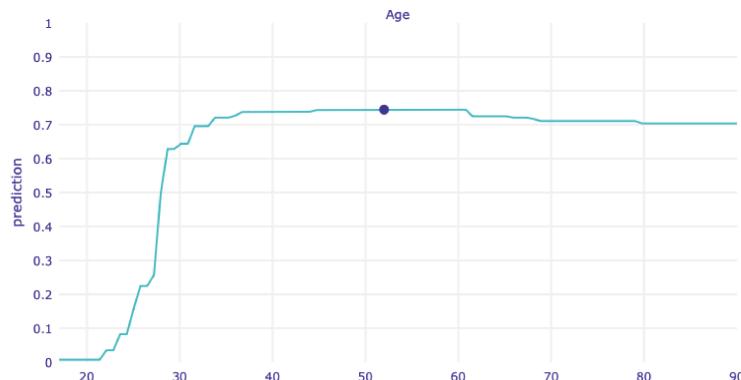
Model Prediction: 0.413



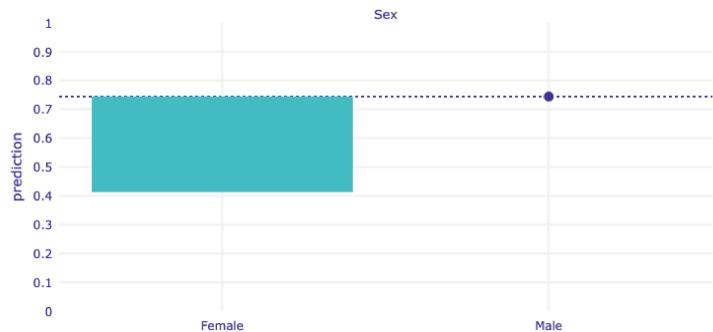
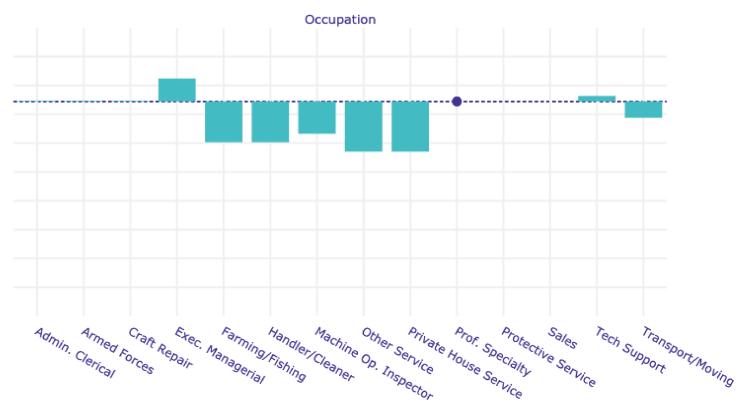
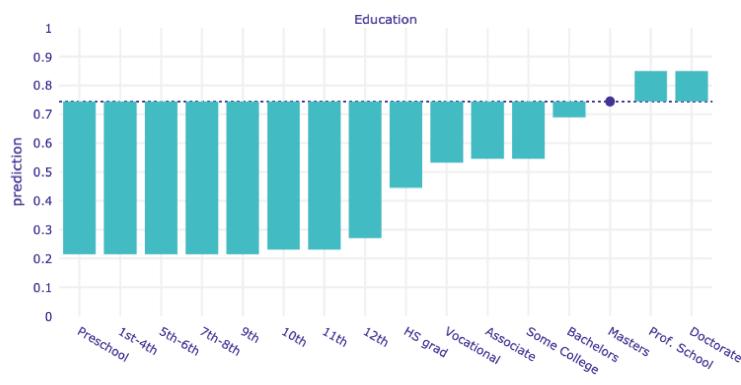
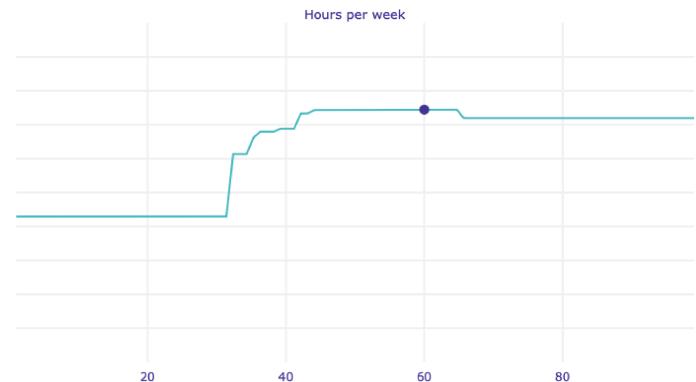
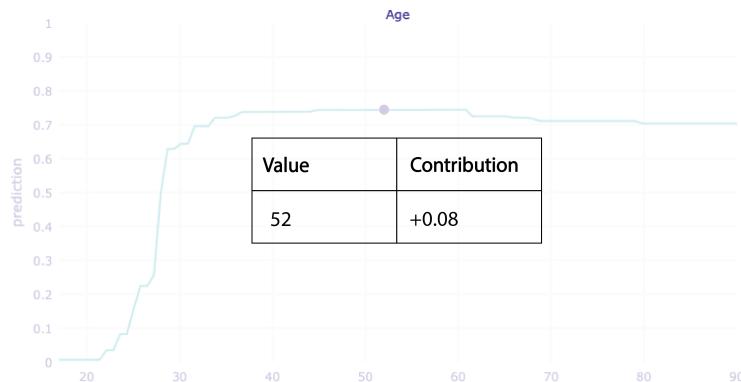
Model Prediction: 0.413



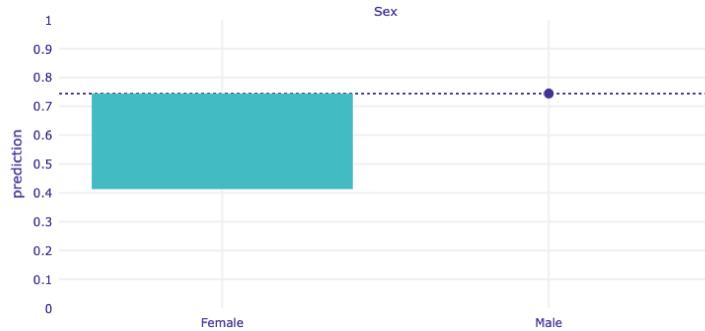
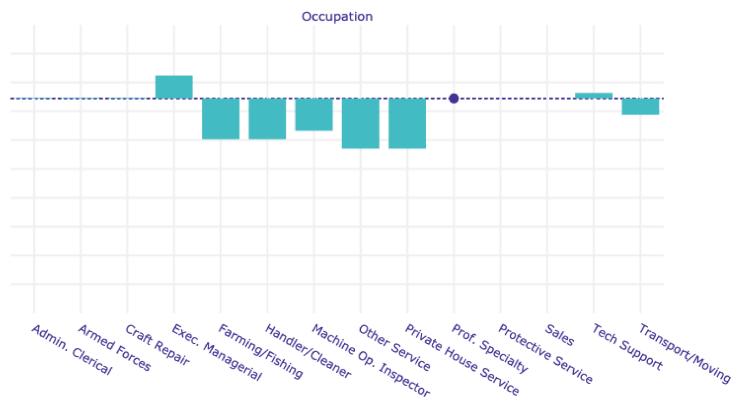
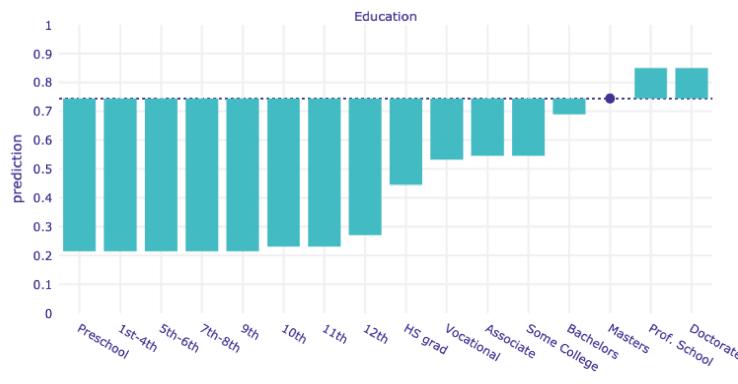
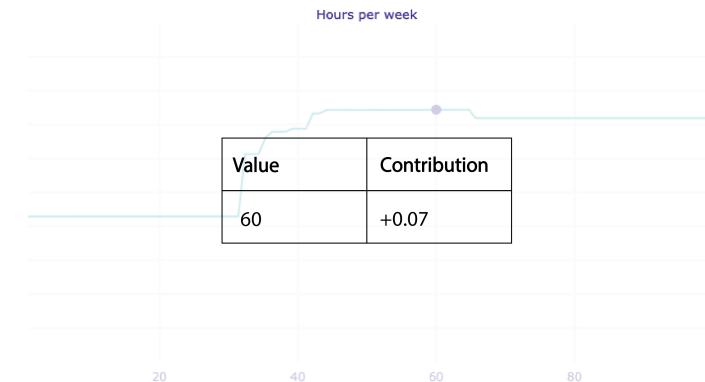
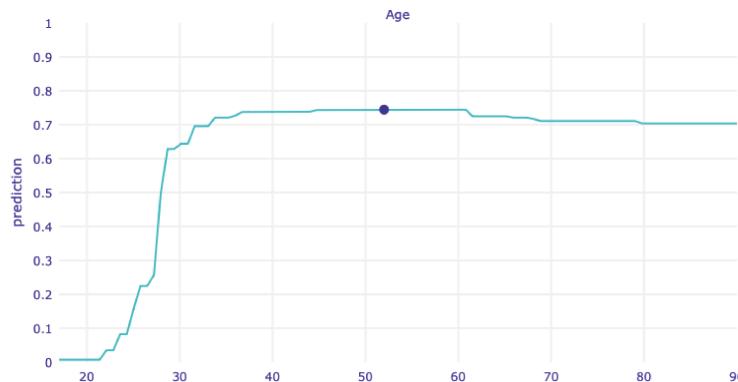
Person B

Model Prediction: 0.745

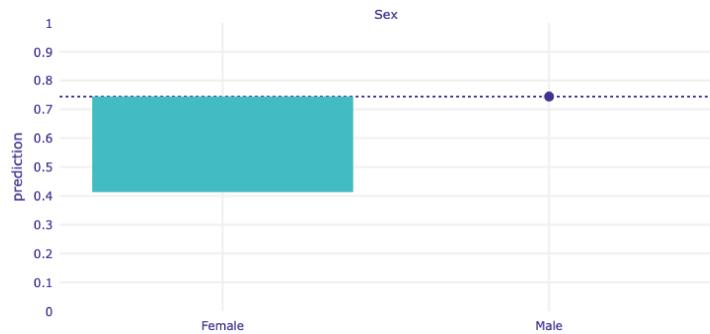
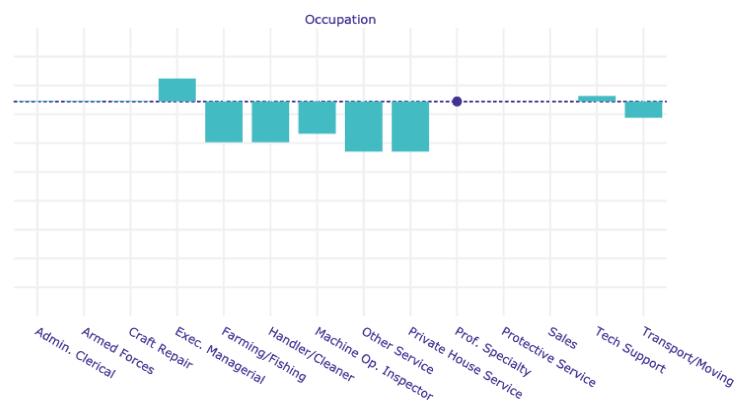
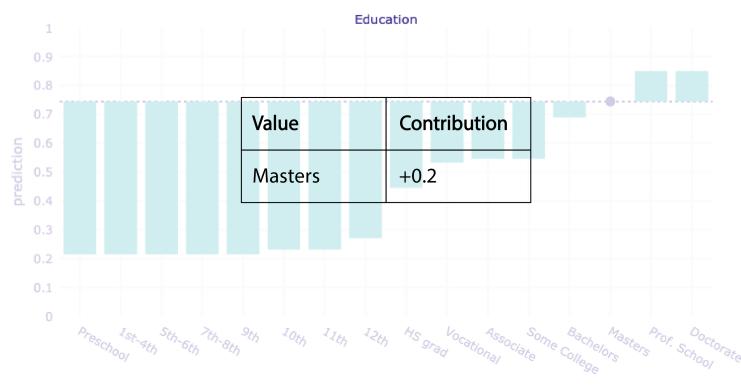
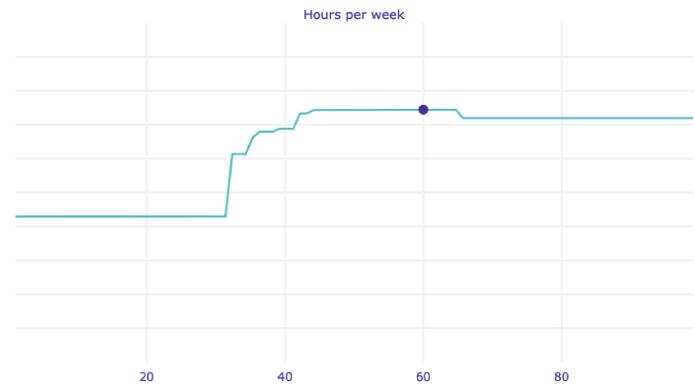
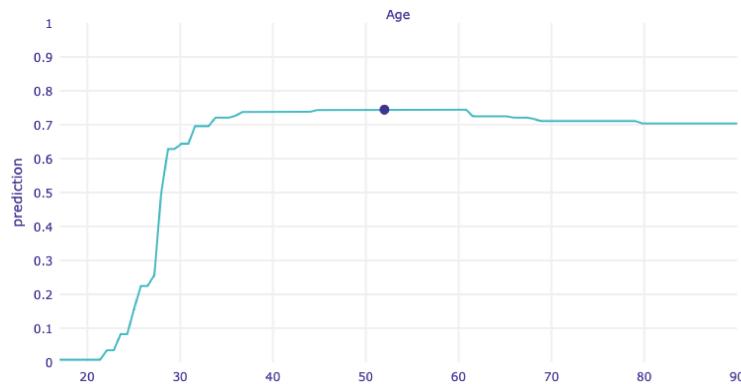
Model Prediction: 0.745



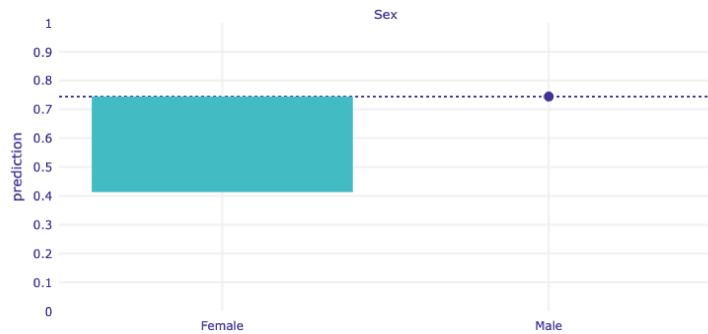
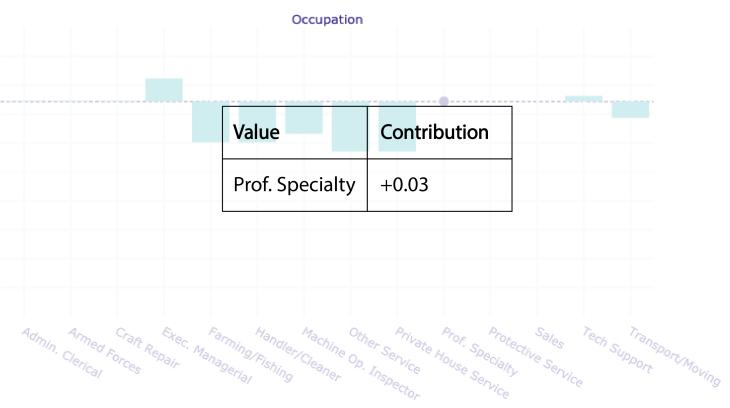
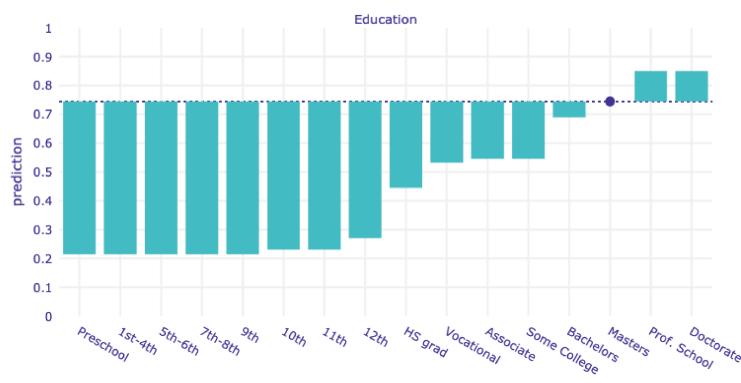
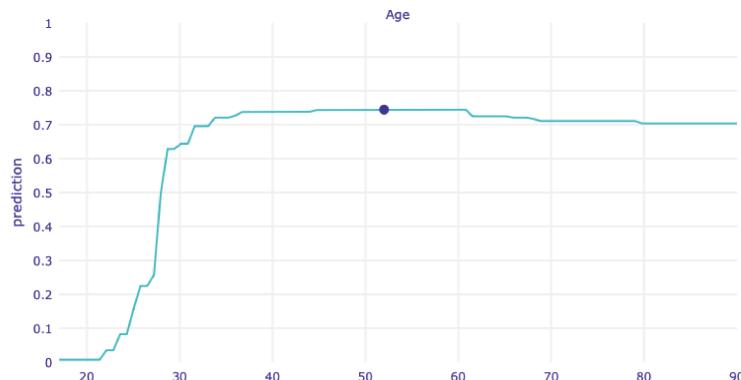
Model Prediction: 0.745



Model Prediction: 0.745



Model Prediction: 0.745



Model Prediction: 0.745



When answering the previous questions about fairness, which design aspects of the given visualizations did you find **most** useful?

When answering the previous questions about fairness, which design aspects of the given visualizations did you find **least** useful?

Demographics

What is your age? Please enter a number.

What is your gender?

- Man/Male (Cis or Trans)
- Woman/Female (Cis or Trans)
- Non-binary

My Gender is Not Listed Above: (Open Text Box)

Unsure/Questioning

Prefer Not to Answer

What is your race/ethnicity?

White

Black/African American

Hispanic/Latinx

Asian

Native American

Hawaiin/Pacific Islander

Other

How much is your yearly income?

\$0 - \$49,999

\$50,000 - \$99,999

\$100,000+

Other

What is the highest level of school you have completed or the highest degree you have received?

- Less than high school degree
- High school graduate (high school diploma or equivalent including GED)
- Some college but no degree
- Associate degree in college (2-year)
- Bachelor's degree in college (4-year)
- Master's degree
- Professional degree (JD, MD, PhD)
- Prefer not to answer

What is your familiarity with machine learning models?

- No familiarity
- Beginner
- Intermediate
- Expert

Feedback

Please give any feedback or suggestions you may have about

this survey

Powered by Qualtrics