0.0.1 Question 2c: Verify Outcome

Did the candidate win or lose the election? Verify with election outcome.

The candidate with the largest increase won the election. The candidate with the largest decrease lost the election. The election outcomes are shown below.

```
In [16]: risingwin = election_sub.query("candidate=='Sharice Davids'").query("forecast_date=='2018-11-0
         fallingwin = election_sub.query("candidate=='Kevin Yoder'").query("forecast_date=='2018-11-06'
         print("Rising candidate: ", rising_candidate, " / Outcome: ", int(risingwin))
         print("Falling candidate: ", falling_candidate, " / Outcome: ", int(fallingwin))
```

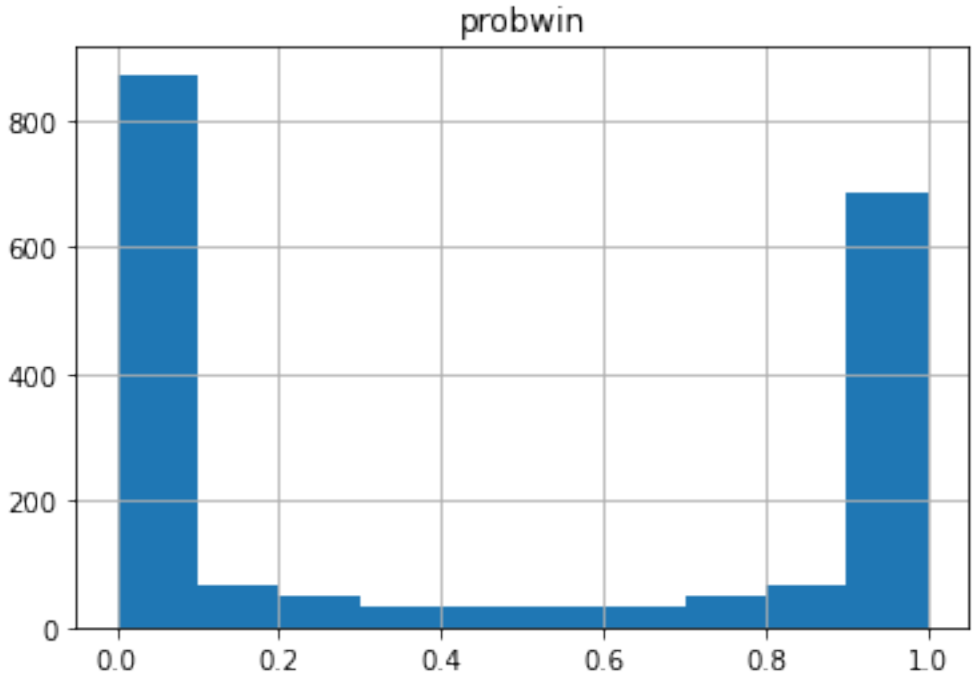
```
Rising candidate:  Sharice Davids  / Outcome:  1
Falling candidate:  Kevin Yoder  / Outcome:  0
```

0.0.2 Question 3a: Prediction Histogram

Make a histogram showing the predicted win probabilities *on the morning of the election.* Again, restrict yourself to only the `classic` predictions.

```
In [17]: election_sub.query("forecast_type == 'classic'").hist(column='probwin')
```

```
Out[17]: array([[<AxesSubplot:title={'center':'probwin'}>]], dtype=object)
```



0.0.3 Question 3b: Prediction difficulty

Are most house elections easy to forecast or hard to forecast? State your reasoning.

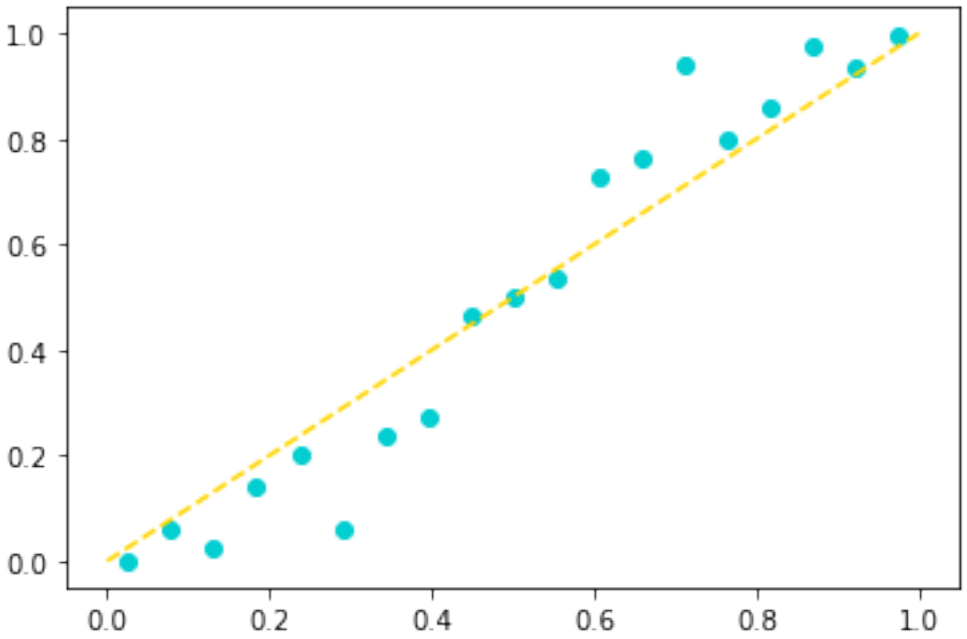
I would say most house elections are hard to forecast, as various factors go into a candidate's popularity and all have to be accounted for. The polls can shift drastically day by day depending on a candidate's choices as well.

0.0.4 Question 4c: Visualize Results

Now make a scatterplot using `midpoints` as the x variable and `fraction_outcome` as the y variable. Draw a dashed line from `[0,0]` to `[1,1]` to mark the line y=x.

In [22]:
```
# magic for showing figures inline
%matplotlib inline
import matplotlib.pyplot as plt
x = np.linspace(0, 1, 10)
y = x
plt.scatter(midpoints, fraction_outcome, color='darkturquoise')
plt.plot(x, y, linestyle='dashed', color='gold')
```

Out[22]:
```
[<matplotlib.lines.Line2D at 0x7f6244160ac0>]
```

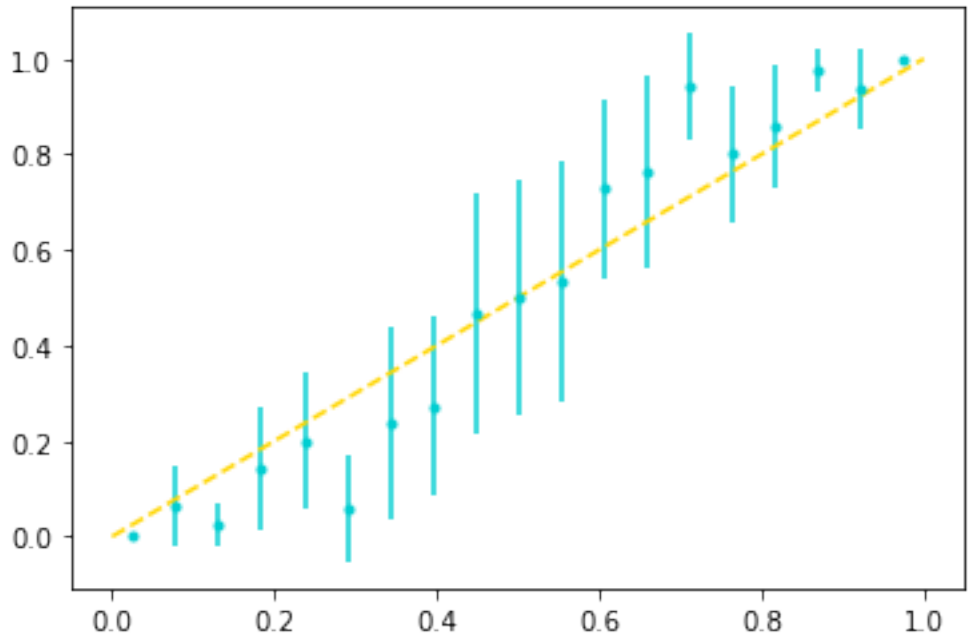


0.0.5 Question 5b: Visualize Error Bars 1

Use `plt.errorbar` to create a new plot with error bars associated with the actual fraction of wins in each bin. Again add a dashed y=x line. Set the argument `fmt='.'` to create a scatterplot with errorbars.

In [25]:
```
# Plotting code below
plt.errorbar(midpoints, fraction_outcome.values, color='darkturquoise', fmt='.', yerr=election_
plt.plot(x, y, linestyle='dashed', color='gold')
```

Out[25]:
```
[<matplotlib.lines.Line2D at 0x7f62440ff0d0>]
```



0.0.6 Question 5d: Understanding Confidence Intervals

Are the 95% confidence intervals generally larger or smaller for more confident predictions (e.g. the predictions closer to 0 or 1). What are the factors that determine the length of the confidence intervals?

The confidence intervals are overall smaller for more confident predictions. Larger sample sizes decrease length of confidence intervals as they can be constructed on more data. Low variability also decreases the length of the confidence intervals. On the other hand, high variability or small sample size means that confidence intervals will cover a larger segment of data.

0.0.7 (PSTAT 234) Question 5f. Visualize Error Bars 2

By now, we have a distribution of success probabilities saved in `bootstrap_election_agg`. We can compute empirical error bars from 2.5% and 97.5% quantiles. Write function named `bootstrap_error_bars` that can be used to calculate the following columns:

- `mean`: mean of probabilities of success
- `err_low`: low point of the error bars
- `err_high`: high point of the error bars

Funtion `bootstrap_error_bars` is to be called by using `bootstrap_election_100_agg.apply(bootstrap_error_bars, ...)`.

```
In [30]: def bootstrap_error_bars(x):
             out = pd.Series([x.mean(), x.mean()-x.quantile(0.025), x.quantile(0.975)-x.mean()],
                             index=['mean', 'err_low', 'err_high'])
             return(out)
```

0.0.8 (PSTAT 234) Question 5g: Interpreting the Results

Are the 95% confidence intervals generally larger or smaller for more confident predictions (e.g. the predictions closer to 0 or 1). What are the factors that determine the length of the error bars?

Compare and contrast model-based error bars and empirically obtained error bars. What are the advantages and disadvantages of these two approaches?

Type your answer here, replacing this text.