In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.utils import shuffle
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import train_test_split

%matplotlib inline
```

In [2]:
```
df = pd.read_csv("Advertising.csv")
```

In [3]:
```
df.head()
```

Out[3]:

	TV	Radio	Newspaper	Sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

In [4]:
```
x = df[["TV"]]
y = df["Sales"]
```

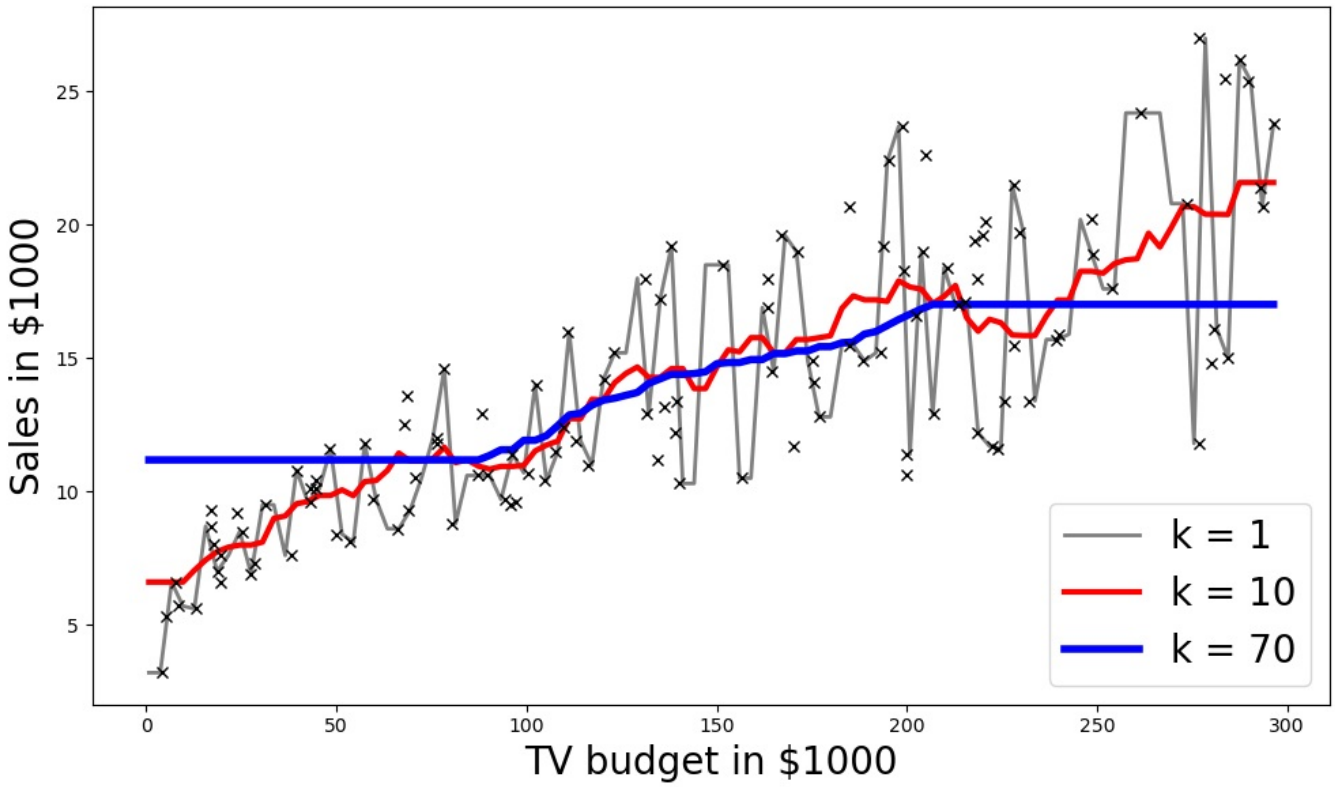
In [5]:
```
x_train, x_test, y_train, y_test = train_test_split(
    x, y, train_size=0.6, random_state=66
)
```

In [6]:
```
k_value_min = 1
k_value_max = 70
k_list = np.linspace(k_value_min, k_value_max, num=70, dtype=int)
```

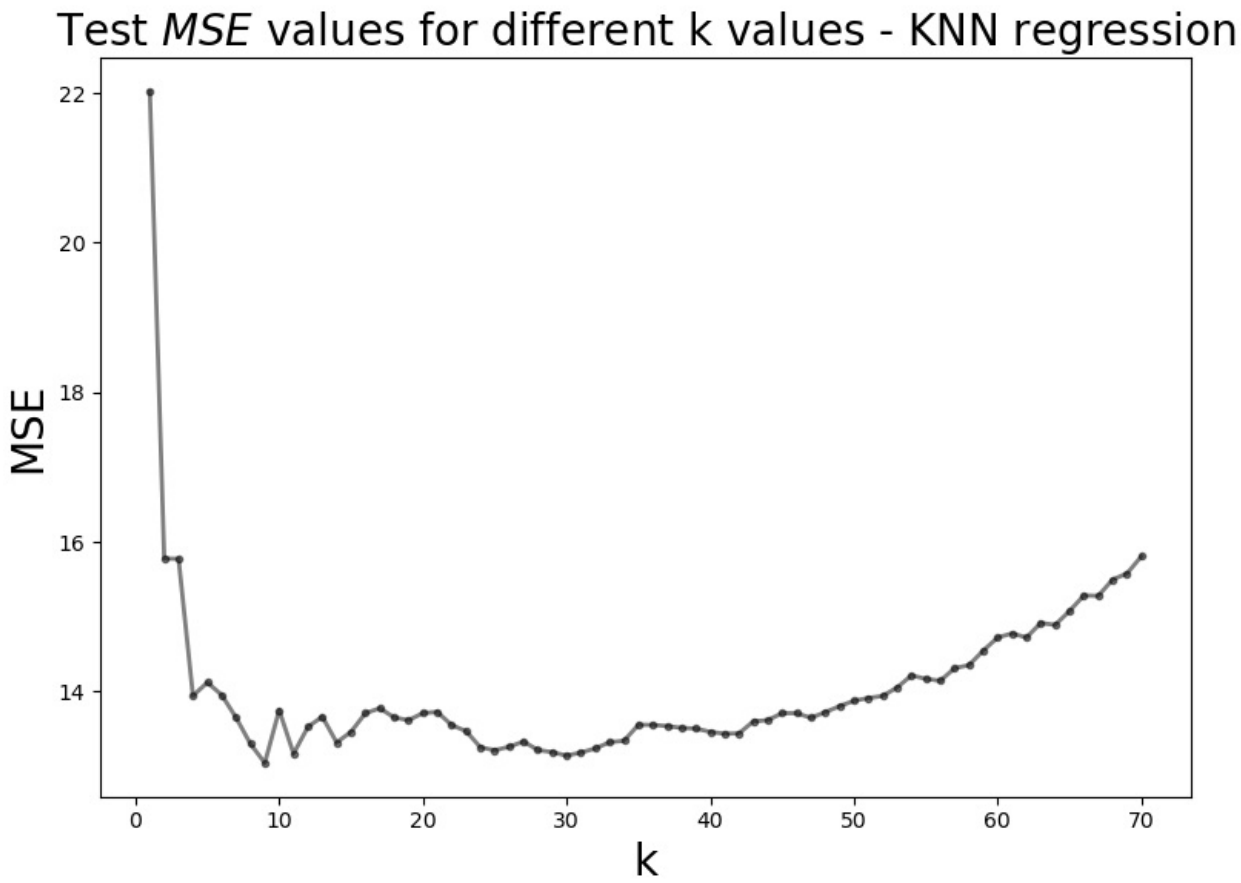
In [7]:
```
fig, ax = plt.subplots(figsize=(10, 6))

knn_dict = {
    1: 22.017374999999998,
    2: 15.7804375,
    3: 15.774361111111114,
    4: 13.9448671875,
    5: 14.122969999999999,
    6: 13.956041666666664,
    7: 13.653349489795914,
    8: 13.302406249999999,
    9: 13.046766975308643,
    10: 13.7528575,
    11: 13.171690082644627,
    12: 13.52842621527778,
    13: 13.66633801775148,
    14: 13.320392219387752,
    15: 13.463508333333333,
    16: 13.712310058593749,
    17: 13.778170847750863,
    18: 13.655200231481482,
    19: 13.618996537396123,
    20: 13.713396875,
    21: 13.729553571428568,
    22: 13.553932076446278,
    23: 13.47318974480151,
    24: 13.25764474826389,
    25: 13.216735400000001,
    26: 13.263455066568048,
    27: 13.333654835390945,
    28: 13.22022337372449,
    29: 13.191935196195008,
    30: 13.145377916666666,
    31: 13.187034079084288,
    32: 13.241443969726564,
    33: 13.324962121212124,
    34: 13.344362889273356,
    35: 13.557367448979594,
    36: 13.556133391203707,
    37: 13.54022854273192,
    38: 13.515847731994464,
    39: 13.507450115055889,
    40: 13.462729140625001,
    41: 13.438169541939322,
    42: 13.442862103174601,
    43: 13.603524405083828,
    44: 13.619427944214873,
```

```
    45: 13.713009444444447,
    46: 13.7120730741966,
    47: 13.650803983703032,
    48: 13.725627658420137,
    49: 13.80626692003332,
    50: 13.878447849999997,
    51: 13.914827374086892,
    52: 13.94275300480769,
    53: 14.055916073335705,
    54: 14.216514188957476,
    55: 14.17145338842975,
    56: 14.144861447704079,
    57: 14.313715912588489,
    58: 14.354166171224733,
    59: 14.549428145647806,
    60: 14.724215972222225,
    61: 14.77866826793873,
    62: 14.723883812434963,
    63: 14.918176744771984,
    64: 14.892030364990237,
    65: 15.07816426035503,
    66: 15.285949408861338,
    67: 15.28058554243707,
    68: 15.498347318339096,
    69: 15.579735586011344,
    70: 15.807493367346936,
}
new_k_list = knn_dict.items()
new_k_list = sorted(new_k_list)
X, Y = zip(*new_k_list)

j = 0

for k_value in k_list:

    model = KNeighborsRegressor(n_neighbors=int(k_value))

    model.fit(x_train, y_train)

    y_pred = model.predict(x_test)
    MSE = mean_squared_error(y_test, y_pred)

    knn_dict[k_value] = MSE

    colors = ["grey", "r", "b"]
    if k_value in [1, 10, 70]:
        xvals = np.linspace(x.min(), x.max(), 100)
        ypreds = model.predict(xvals)
        ax.plot(
            xvals,
            ypreds,
            "-",
            label=f"k = {int(k_value)}",
            linewidth=j + 2,
            color=colors[j],
        )
        j += 1

ax.legend(loc="lower right", fontsize=20)
ax.plot(x_train, y_train, "x", label="test", color="k")
ax.set_xlabel("TV budget in $1000", fontsize=20)
ax.set_ylabel("Sales in $1000", fontsize=20)
plt.tight_layout()
```

```
C:\Users\cowbo\Anaconda\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names
, but KNeighborsRegressor was fitted with feature names
  warnings.warn(
C:\Users\cowbo\Anaconda\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names
, but KNeighborsRegressor was fitted with feature names
  warnings.warn(
C:\Users\cowbo\Anaconda\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names
, but KNeighborsRegressor was fitted with feature names
  warnings.warn(
```



In [8]:
```
plt.figure(figsize=(8, 6))
plt.plot(X, Y, "k.-", alpha=0.5, linewidth=2)
plt.xlabel("k", fontsize=20)
plt.ylabel("MSE", fontsize=20)
plt.title("Test $MSE$ values for different k values - KNN regression", fontsize=20)
plt.tight_layout()
```



In [9]:
```
min_mse = min(Y)

best_model = [key for (key, value) in knn_dict.items() if value == min_mse]

print("The best k value is ", best_model, "with a MSE of ", min_mse)
```

```
The best k value is  [9] with a MSE of  13.046766975308643
```

In [10]:
```
model = KNeighborsRegressor(n_neighbors=best_model[0])
model.fit(x_train, y_train)
y_pred_test = model.predict(x_test)

print(f"The R2 score for your model is {r2_score(y_test, y_pred_test)}")
```

```
The R2 score for your model is 0.5492457002030715
```