Explanation:

Percent salary hike represents the rate of increase of employee wages over a fixed period of time. It reflects the company's financial and performance performance, and it also depends on the individual performance and working conditions of the employees. Employees that have a good PerformanceRating(>=3), good education(>=3), good JobInvolvement(>=3) and good job level(>=3) but still have a PercentSalaryHike lower than average is a business success.

There are five people who are business failures. A notable feature is that their years of work and the time they follow the current manager are relatively long, which means you are likely to receive above-average salary growth because of your "loyalty". There are a total 55 people who are business success whose "PercentSalaryHike"<15.21 and "PerformanceRating" >2 and "Education">3 and "JobInvolvement">2 and "JobLevel">2 which shows that Master & Phds that are under-rewarded.

Department	EducationField	EnvironmentSatisfaction	MaritalStatus	TotalWorkingYears	StockOptionLevel	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager	TrainingTimesLastYea
text	text	int8	text	int8	int8	int8	int8	int8	int8
Research & Development	Technical Degree	3	Single	4	0	2	1	2	3
Research & Development	Life Sciences	1	Married	6	0	2	3	4	2
Sales	Life Sciences	2	Single	3	0	2	2	2	2
Research & Development	Life Sciences	1	Married	6	0	3	0	4	2
Research & Development	Life Sciences	2	Married	6	1	4	1	3	2

The average salary hike is 15.21, but the research department has an average hike of 15.29% while the HR department only has 14.76%. It means if you want a higher increase in salary, don’t work as a HR. Also, Research & Development has a total of 369 that are higher than average in salary hike compared to only 21 people in the Sales Department. In a word, loyalty & department > Work you done & Degree.

Code example:

```
SELECT
"Department","EducationField","EnvironmentSatisfaction","MaritalStatus","TotalWorkingYears"
,"StockOptionLevel","YearsInCurrentRole","YearsSinceLastPromotion","YearsWithCurrManage
r","TrainingTimesLastYear"
FROM "DATA_ibm_mongo_prepared"
Where "PercentSalaryHike">15.21 and "PerformanceRating" <4 and "Education"<2 and
"JobInvolvement"<3 and "JobLevel"<2

SELECT *
FROM "DATA_ibm_mongo_prepared"
Where "PercentSalaryHike">15.21

SELECT
"Department","EducationField","EnvironmentSatisfaction","MaritalStatus","TotalWorkingYears"
,"StockOptionLevel","YearsInCurrentRole","YearsSinceLastPromotion","YearsWithCurrManage
r","TrainingTimesLastYear"
FROM "DATA_ibm_mongo_prepared"
Where "PercentSalaryHike"<15.21 and "PerformanceRating" >2 and "Education">3 and
"JobInvolvement">2 and "JobLevel">2

SELECT "Department", count("PercentSalaryHike")
```

```
 FROM "DATA_ibm_mongo_prepared"
   Where "PercentSalaryHike">15.21
     Group by "Department"
```

Department text	count int8
Research & Development	369
Sales	161
Human Resources	21

```
SELECT  AVG("PercentSalaryHike") as Avg_Hike
 FROM "DATA_ibm_mongo_prepared"
```

avg_hike numeric
15.209523809523809

```
SELECT "Department", AVG("PercentSalaryHike") as Avg_Hike
 FROM "DATA_ibm_mongo_prepared"
   Group by "Department"
```

Department text	avg_hike numeric
Sales	15.096412556053812
Research & Development	15.291363163371487
Human Resources	14.761904761904763

2A

42 Pokemon are Flying or Grass (in Type_1) and have an Attack greater than 65. We can interpret the result by counting the conditions that Attack less than 65, and only Flying or Grass in Type_1 that have an Attack greater than 65. There are 74 Pokemon which are either flying or grass type so more than half of them have an attack higher than 65 which is nor bad. Also, flying in Type_1 that have an Attack less than 65 have only 3 Pokemon, which means there are only 3 flying pokemon that have a high attack compared to Grass type, so when you want to catch an attacker, you would consider a Grass type rather than a flying type.

The average Speed of Pokemon that have Attack greater than 75 and are non-Legendary is 72.80060422960725. We can interpret by comparing them with the condition that is Legendary true. The average Speed of all the Pokemon that have Attack greater than 75 and are Legendary is 102.36666666666666, which means that Legendary pokemon are more likely to have a high speed than non-legendary pokemon when they are having a relatively high basic attack. In addition, the speed of non-legendary pokemon that have Attack less than 75 is 59.165333333333336 which means lower attacks are usually with low speed.

Code continue:

```
[{$match: {
  "Type 1":{$in:["Flying","Grass"]},
  "Attack":{$gt:65}
}}, {$group: {
```

```
  _id: "Flying or Grass and have Attack greater than 65",
  count: { $sum: 1 }
  }}]
```

And:

```
[{$match: {
  "Type 1":{$in:["Flying","Grass"]},

  }}, {$group: {
  _id: "Flying or Grass and have Attack greater than 65",
  count: { $sum: 1 }
  }}]
```

And:

```
[{$match: {
  "Type 1":{$in:["Flying"]},

  }}, {$group: {
  _id: "Flying or Grass and have Attack greater than 65",
  count: { $sum: 1 }
  }}]

# continue
[{$match: {
```

```
  Legendary:false,
  Attack:{$gt:75}
}}, {$group: {
  _id: "pokemon_speed",
  avg_speed: {
    $avg: "$Speed"}
}}, {}]
```

And:

```
[{$match: {
  Legendary: true,
  Attack:{$gt:75}
}}, {$group: {
  _id: "pokemon_speed",
  avg_speed: {
    $avg: "$Speed"}
}}, {}]
```

And:

```
[{$match: {
  Legendary: false,
  Attack:{$lt:75}
}}, {$group: {
  _id: "pokemon_speed",
```

```
  avg_speed: {
    $avg: "$Speed"}
}}, {}]
```

3A

```
Populating the interactive namespace from numpy and matplotlib

+--------+-----+-------------------+-------------------+----------+
|features|label|      rawPrediction|        probability|prediction|
+--------+-----+-------------------+-------------------+----------+
|  [65.0]|    0|[3.86994840894393...|[0.97956678008021...|       0.0|
|  [80.0]|    0|[3.12215447116326...|[0.95779740114518...|       0.0|
| [100.0]|    0|[2.12509588745570...|[0.89331854452362...|       0.0|
| [122.0]|    0|[1.02833144537738...|[0.73659228386557...|       0.0|
|  [60.0]|    0|[4.11921305487082...|[0.98400276881993...|       0.0|
|  [80.0]|    0|[3.12215447116326...|[0.95779740114518...|       0.0|
| [109.0]|    0|[1.67641952478730...|[0.84242983487456...|       0.0|
| [130.0]|    0|[0.62950801189436...|[0.65237789701580...|       0.0|
| [159.0]|    0|[-0.8162269344815...|[0.30656516342118...|       1.0|
|  [50.0]|    0|[4.61774234672460...|[0.99022149795980...|       0.0|
|  [65.0]|    0|[3.86994840894393...|[0.97956678008021...|       0.0|
|  [85.0]|    0|[2.87288982523637...|[0.94648989729889...|       0.0|
| [135.0]|    0|[0.38024336596747...|[0.59393179850507...|       0.0|
|  [20.0]|    0|[6.11333022228594...|[0.99779171911228...|       0.0|
```

```
|  [25.0]|    0|[5.86406557635905...|[0.99716836565616...|       0.0|
|  [90.0]|    0|[2.62362517930948...|[0.93236666546768...|       0.0|
|  [20.0]|    0|[6.11333022228594...|[0.99779171911228...|       0.0|
|  [25.0]|    0|[5.86406557635905...|[0.99716836565616...|       0.0|
|  [45.0]|    0|[4.86700699265149...|[0.99236241551627...|       0.0|
|  [15.0]|    0|[6.36259486821283...|[0.99827808505502...|       0.0|
+--------+-----+--------------------+--------------------+----------+
only showing top 20 rows


AUC ROC:0.893354264783
('Model Intercept: ', -2.4254832293172)
+--------------------+
|     Feature Weight|
+--------------------+
|0.00536437531740333|
+--------------------+


+-----+----------+--------------------+--------+
|label|prediction|         probability|features|
+-----+----------+--------------------+--------+
|    0|       0.0|[0.88862761232493...|  [65.0]|
|    0|       0.0|[0.88041155002828...|  [80.0]|
|    0|       0.0|[0.86864668331565...| [100.0]|
|    0|       0.0|[0.85458564533941...| [122.0]|
|    0|       0.0|[0.89125459491084...|  [60.0]|
```

```
|  0|      0.0|[0.88041155002828...|  [80.0]|
|  0|      0.0|[0.86303931350076...| [109.0]|
|  0|      0.0|[0.84917107727992...| [130.0]|
|  0|      0.0|[0.82814647899864...| [159.0]|
|  0|      0.0|[0.89634566495260...|  [50.0]|
|  0|      0.0|[0.88862761232493...|  [65.0]|
|  0|      0.0|[0.87755861523467...|  [85.0]|
|  0|      0.0|[0.84570347548523...| [135.0]|
|  0|      0.0|[0.91037283119846...|  [20.0]|
|  0|      0.0|[0.90816010045524...|  [25.0]|
|  0|      0.0|[0.87464731046284...|  [90.0]|
|  0|      0.0|[0.91037283119846...|  [20.0]|
|  0|      0.0|[0.90816010045524...|  [25.0]|
|  0|      0.0|[0.89881132892405...|  [45.0]|
|  0|      0.0|[0.91253738420155...|  [15.0]|
+-----+----------+--------------------+--------+
only showing top 20 rows
```

3B

Business insight from the model: The quality of this model is Excellent & outstanding discrimination because it has an AUC ROC:0.893354264783 which is very close to 0.9, An AUC ROC that is more close to one is considered to be perfect, and those below 0.5 would be considered trash. Therefore, the model is not perfect but it is statistically Excellent and super close to outstanding level.

The B1 of this model is 0.00536437531740333. It is shown as Feature Weight in the table above. The value is positive which means an increase in Special attack, the probability of the legendary true (business success) would go up. It means that whenever a Special attack increase by 1 unit, the odds would increase by exp(0.00536437531740333) which is 0.00537878934.