

Dokumentace 2. části projektu do předmětu UPA

Žaneta Grossová(xgross11)
David Hurta(xhurta04)

zima 2022

Obsah

| | | |
|-----------|---|-----------|
| I | Explorativní analýza zvolené datové sady | 1 |
| 1 | Atributy datové sady | 2 |
| 2 | Rozložení hodnot jednotlivých atributů | 8 |
| 3 | Odlehlé hodnoty | 16 |
| 4 | Analýza chybějících hodnot | 19 |
| 5 | Korelační analýza numerických atributů | 21 |
| | | |
| II | Datové sady vhodné pro dolovací algoritmy | 23 |
| 6 | Odstranění irelevantních atributů | 24 |
| 7 | Zpracování chybějících hodnot | 25 |
| 8 | Zpracování odlehlých hodnot | 26 |
| 9 | Datová sada číslo I. | 27 |
| 9.1 | Diskretizace numerických atributů | 27 |
| 10 | Datová sada číslo II. | 29 |
| 10.1 | Transformace kategorických atributů na numerické a jejich následná normalizace | 29 |

Část I

Explorativní analýza zvolené datové sady

Kapitola 1

Atributy datové sady

Tato kapitola popisuje jednotlivé atributy vyskytující se v datové sadě Tučňáci¹. Podrobně popisuje typy a hodnoty, kterých můžou atributy nabývat. Samotná datová sada se skládá z následujících 17 atributů.

Atribut `studyName`

Tento atribut popisuje název studie, pod kterým byl záznam přidán do datové sady.

| Datový typ | Celkový počet | Počet unikátních hodnot | Nejčastější hodnota |
|------------|---------------|-------------------------|---------------------|
| Řetězec | 344 | 3 | PAL0910 |

Tabulka 1.1: Charakteristika atributu *studyName*.

| Hodnota | Celkový počet |
|---------|---------------|
| PAL0910 | 120 |
| PAL0809 | 114 |
| PAL0708 | 110 |

Tabulka 1.2: Kategorie atributu *studyName*.

¹Datová sada Palmer Archipelago (Antarctica) penguin data <https://www.kaggle.com/datasets/parulpandey/palmer-archipelago-antarctica-penguin-data>

Atribut Species

Tento atribut popisuje druh tučňáka, který byl přidán do datové sady.

| Datový typ | Celkový počet | Počet unikátních hodnot | Nejčastější hodnota |
|------------|---------------|-------------------------|-------------------------------------|
| Řetězec | 344 | 3 | Adelie Penguin (Pygoscelis adeliae) |

Tabulka 1.3: Charakteristika atributu *Species*.

| Hodnota | Celkový počet |
|---|---------------|
| Adelie Penguin (Pygoscelis adeliae) | 152 |
| Gentoo penguin (Pygoscelis papua) | 124 |
| Chinstrap penguin (Pygoscelis antarctica) | 68 |

Tabulka 1.4: Kategorie atributu *Species*.

Atribut Island

Tento atribut popisuje ostrov, kde byl tučňák zaznamenán do datové sady.

| Datový typ | Celkový počet | Počet unikátních hodnot | Nejčastější hodnota |
|------------|---------------|-------------------------|---------------------|
| Řetězec | 344 | 3 | Biscoe |

Tabulka 1.5: Charakteristika atributu *Island*.

| Hodnota | Celkový počet |
|-----------|---------------|
| Biscoe | 168 |
| Dream | 124 |
| Torgersen | 52 |

Tabulka 1.6: Kategorie atributu *Island*.

Atribut Clutch Completion

Tento atribut popisuje, zda studijní hnízdo bylo pozorováno s plnou snůškou, tj. 2 vejci.

| Datový typ | Celkový počet | Počet unikátních hodnot | Nejčastější hodnota |
|------------|---------------|-------------------------|---------------------|
| Řetězec | 344 | 2 | Yes |

Tabulka 1.7: Charakteristika atributu *Clutch Completion*.

| Hodnota | Celkový počet |
|---------|---------------|
| Yes | 308 |
| No | 36 |

Tabulka 1.8: Kategorie atributu *Clutch Completion*.

Atribut Sex

Tento atribut popisuje pohlaví daného tučňáka, který byl zaznamenán do datové sady.

| Datový typ | Celkový počet | Počet unikátních hodnot | Nejčastější hodnota |
|------------|---------------|-------------------------|---------------------|
| Řetězec | 334 | 3 | MALE |

Tabulka 1.9: Charakteristika atributu *Sex*.

| Hodnota | Celkový počet |
|---------|---------------|
| MALE | 168 |
| FEMALE | 165 |
| . | 1 |

Tabulka 1.10: Kategorie atributu *Sex*.

Další kategorické atributy

Tato podkapitola obsahuje další kategorické atributy, které datová sada obsahuje, avšak jejich význam je přímočarý v rámci datové sady.

Atribut Region

Tento atribut je datového typu řetězec a jeho hodnota je vyplněna v každém záznamu stejnou hodnotou a to hodnotou **Anvers**. Tato představuje region, ze kterého byl tučňák zaznamenán.

Atribut Stage

Tento atribut je datového typu řetězec a jeho hodnota je vyplněna v každém záznamu stejnou hodnotou a to hodnotou **Adult**, **1 Egg Stage**. Tato hodnota představuje reprodukční fázi při odběru vzorků.

Atribut Individual ID

Tento atribut je datového typu řetězec. Jeho hodnota identifikuje určitého tučňáka. V datové sadě má celkem 344 výskytů a nachází se v ní 190 jedinečných hodnot, což znamená, že více záznamů má stejnou hodnotu atributu *Individual ID* a to nanejvýš třikrát v této datové sadě.

Atribut Date Egg

Tento atribut reprezentuje datum studie hnízda pozorovaného s 1 vejcem. Je datového typu řetězec ve tvaru MM/DD/YY (například 11/15/07). V datové

sadě má celkem 344 výskytů a z toho je 50 jedinečných. Datумы se nachází v intervalu od roku 2007 do roku 2009.

Atribut Comments

Tento atribut je datového typu řetězec a obsahuje text poskytující další relevantní informace pro daný záznam. Tento atribut není v dané datové sadě příliš relevantní, a to z toho důvodu, že je vyplněn pouze u 26 záznamů, u ostatních je vynechán. Jeho nejčastější hodnota je "Nest never observed with full clutch."

Numerické atributy

V této podkapitole jsou podrobně popsány numerické atributy této datové sady. Souhrn základních hodnot těchto atributů se vyskytuje v tabulce 1.11.

Atribut Sample Number

Tento atribut je datového typu celé číslo a označuje číslo vzorku.

Atribut Culmen Length (mm)

Tento atribut je datového typu `float` a označuje délku zobáku daného tučňáka v milimetrech.

Atribut Culmen Depth (mm)

Tento atribut je datového typu `float` a označuje hloubku zobáku daného tučňáka v milimetrech.

Atribut Flipper Length (mm)

Tento atribut je datového typu `float` a označuje délku ploutve daného tučňáka v milimetrech.

Atribut Body Mass (g)

Tento atribut je datového typu `float` a označuje hmotnost daného tučňáka v gramech.

Atribut Delta 15 N (o/oo)

Tento atribut je datového typu `float` a označuje míru poměru stabilních izotopů $^{13}\text{C}:^{12}\text{C}$.

Atribut Delta 13 C (o/oo)

Tento atribut je datového typu `float` a označuje míru poměru stabilních izotopů $^{13}\text{C}:^{12}\text{C}$.

| | Sample Num- ber | Culmen Length (mm) | Culmen Depth (mm) | Flipper Length (mm) | Body Mass (g) | Delta 15 N (o/oo) | Delta 13 C (o/oo) |
|--------|-----------------------|--------------------------|-------------------------|---------------------------|---------------------|-------------------------|-------------------------|
| Počet | 344 | 342 | 342 | 342 | 342 | 330 | 331 |
| Průměr | 63.2 | 43.9 | 17.2 | 200.9 | 4202 | 8.7 | -25.7 |
| STD | 40.4 | 5.5 | 2.0 | 14.1 | 802 | 0.6 | 0.8 |
| Min. | 1 | 32.1 | 13.1 | 172 | 2700 | 7.6 | -27.0 |
| 25% | 29 | 39.2 | 15.6 | 190 | 3550 | 8.3 | -26.3 |
| 50% | 58 | 44.5 | 17.3 | 197 | 4050 | 8.6 | -25.8 |
| 75% | 95.25 | 48.5 | 18.7 | 213 | 4750 | 9.2 | -25.1 |
| Max. | 151 | 59.6 | 21.5 | 231 | 6300 | 10.0 | -23.8 |

Tabulka 1.11: Souhrn charakteristik numerických atributů.

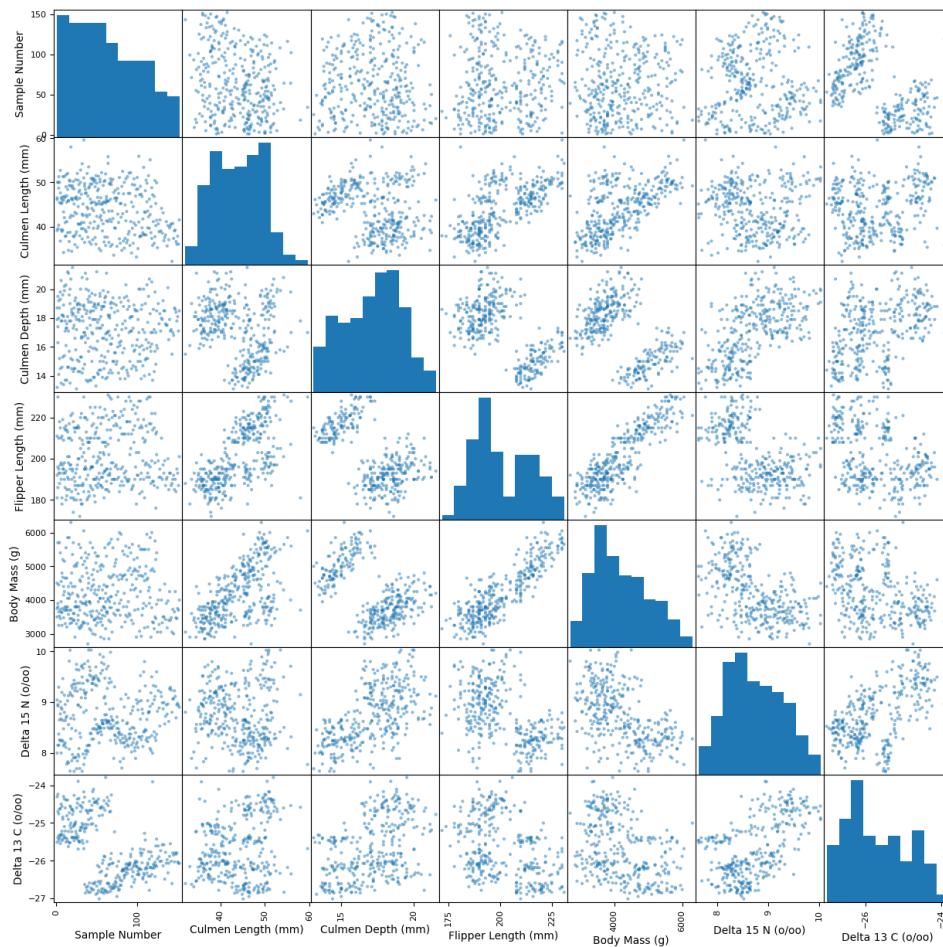
Kapitola 2

Rozložení hodnot jednotlivých atributů

Tato kapitola popisuje prozkoumání rozložení hodnot jednotlivých atributů pomocí vhodných grafů. Veškeré obrázky jsou k dispozici v plné velikosti a kvalitě v přiloženém Jupyter notebooku.

Souhrnné rozložení všech atributů

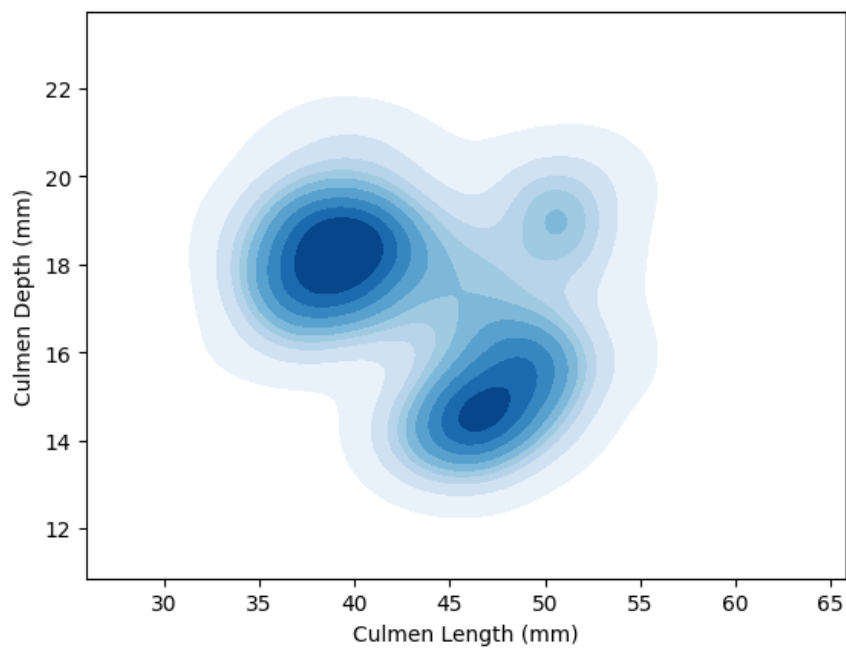
Pro prozkoumání numerických atributů byl použit maticový graf obsahující bodové grafy jednotlivých atributů vůči sobě a obsahující histogramy pro jednotlivé atributy. Graf se vyskytuje na obrázku 2.1. Tento graf poskytuje rychlý přehled vztahů jednotlivých atributů vůči sobě. Z tohoto grafu lze velmi rychle rozpoznat jednoduché vztahy mezi atributy či jejich shlukování.



Obrázek 2.1: Souhrnné rozložení numerických atributů vůči sobě pomocí bodových grafů a histogramů pro každý atribut.

Pravděpodobnostní rozložení atributů zobáku

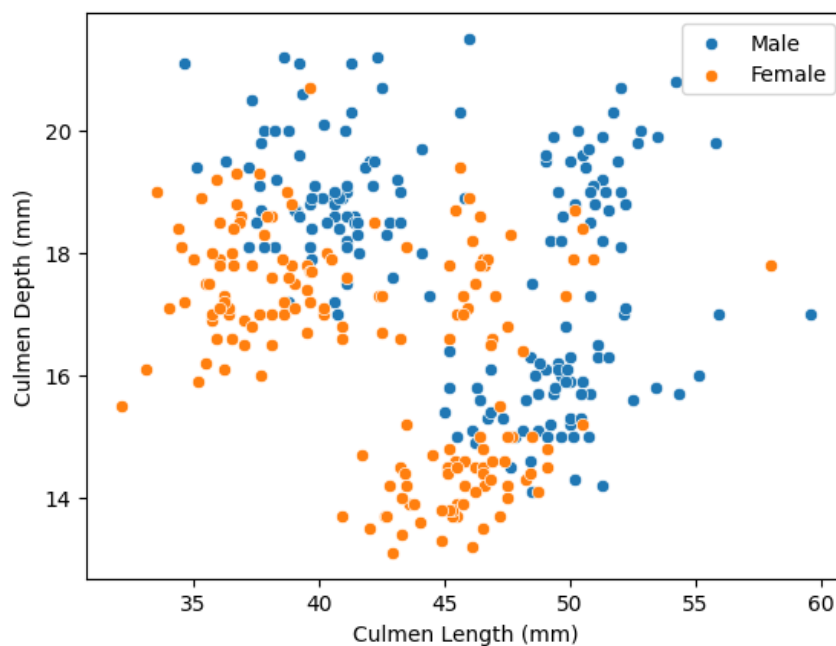
V maticovém grafu 2.1 si můžeme například povšimnout shluků u vztahu mezi délkou a hloubkou zobáku tučňáka. Tento jev byl následně dále prozkoumán pomocí graf popisujícího odhad hustoty pravděpodobnosti metodou KDE (Kernel density estimation). Na grafu si můžeme všimnout tří různých vrcholů, kde nejpravděpodobněji se bude nacházet tato hodnota u tučňáka. Tento graf se nachází na obrázku 2.2.



Obrázek 2.2: Graf popisující odhad hustoty pravděpodobnosti metodou KDE (Kernel density estimation).

Rozložení atributů zobáku vůči pohlaví

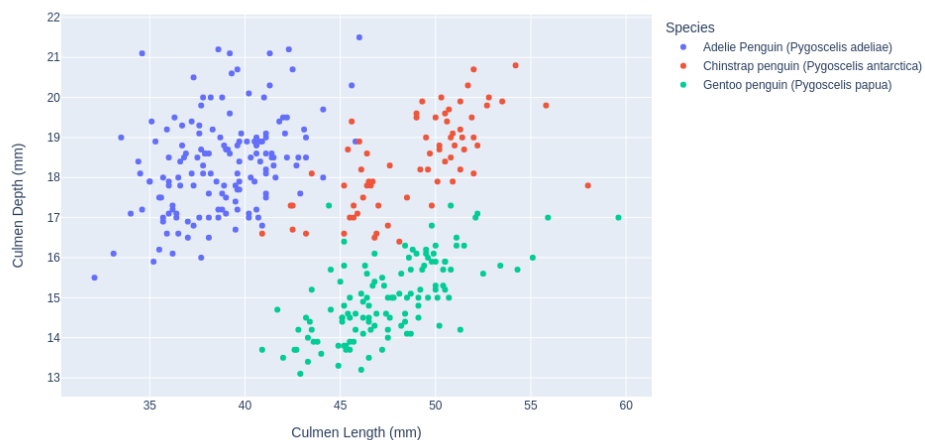
Pro zvýraznění kategorie pohlaví jednotlivých vzorků byl použit bodový graf, který se nachází na obrázku 2.3. Na tomto grafu je barevně zvýrazněno pohlaví tučňáka v jednotlivých bodech. Můžeme si všimnout lineárního posunutí mezi hodnotami jednoho pohlaví v porovnání k druhému pohlaví.



Obrázek 2.3: Bodový graf popisující vztah mezi hloubkou a délkou zobáku tučňáka.

Rozložení atributů zobáku vůči druhu

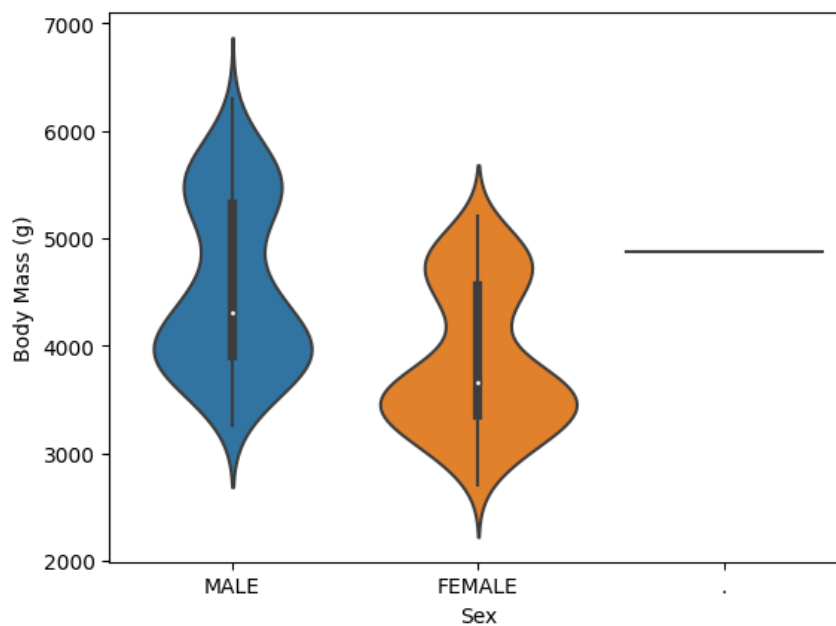
Taktéž si můžeme povšimnout velmi zajímavého vztahu mezi atributy zobáku a druhem daného tučňáka. Tento vztah je znázorněn na bodovém grafu, který se vyskytuje na obrázku 2.4. Z obrázku je zřejmé, že druh tučňáka lze s vysokou jistotou určit pomocí hodnot atributů jeho zobáku, jelikož na grafu se vyskytují znatelně oddělené shluky.



Obrázek 2.4: Bodový graf popisující vztah mezi zobákem a druhem tučňáka.

Rozložení mezi váhou a pohlaví

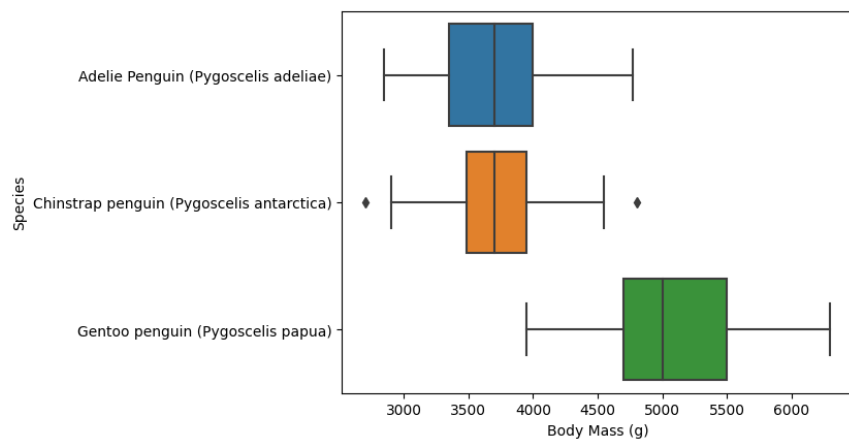
K prozkoumání vztahu mezi váhou a pohlaví tučňáka byl použit houslový graf, který velmi dobře vizuálně znázorňuje tento vztah. Tento graf se vyskytuje na obrázku 2.5. Můžeme zde jasně vidět, že samci tučňáka bývají těžší.



Obrázek 2.5: Houslový graf popisující vztah mezi váhou a pohlaví tučňáka.

Rozložení mezi váhou a druhem

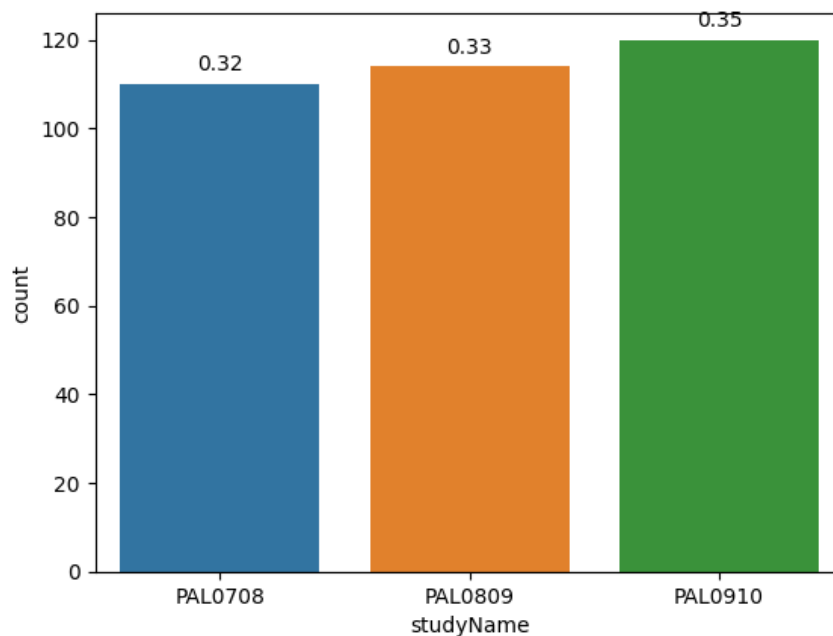
Pro znázornění vztahu mezi váhou a druhem tučňáka byl použit krabicový graf, který se vyskytuje na obrázku 2.6. Při bližším pozorování si můžeme všimnout, že tučňáci druhu Gentoo bývají znatelně těžší než tučňáci jiných druhů.



Obrázek 2.6: Krabicový graf popisující vztah mezi druhem a váhou tučňáka.

Rozložení kategorického atributu typu studie

Graf popisující počet vzorků z jednotlivých studií a jeho celkové procentuální zastoupení se nachází na obrázku 2.7. S menší odchylkou je rozložení rovnoměrné.

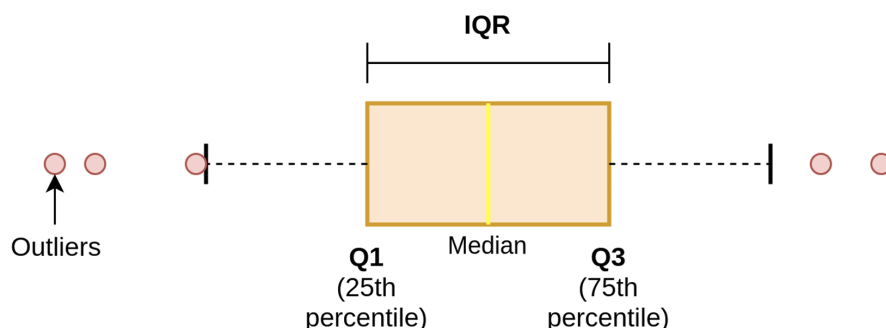


Obrázek 2.7: Početní graf znázorňující počet vzorků u každé kategorie atributu *studyName* a jejich procentuální zastoupení.

Kapitola 3

Odlehlé hodnoty

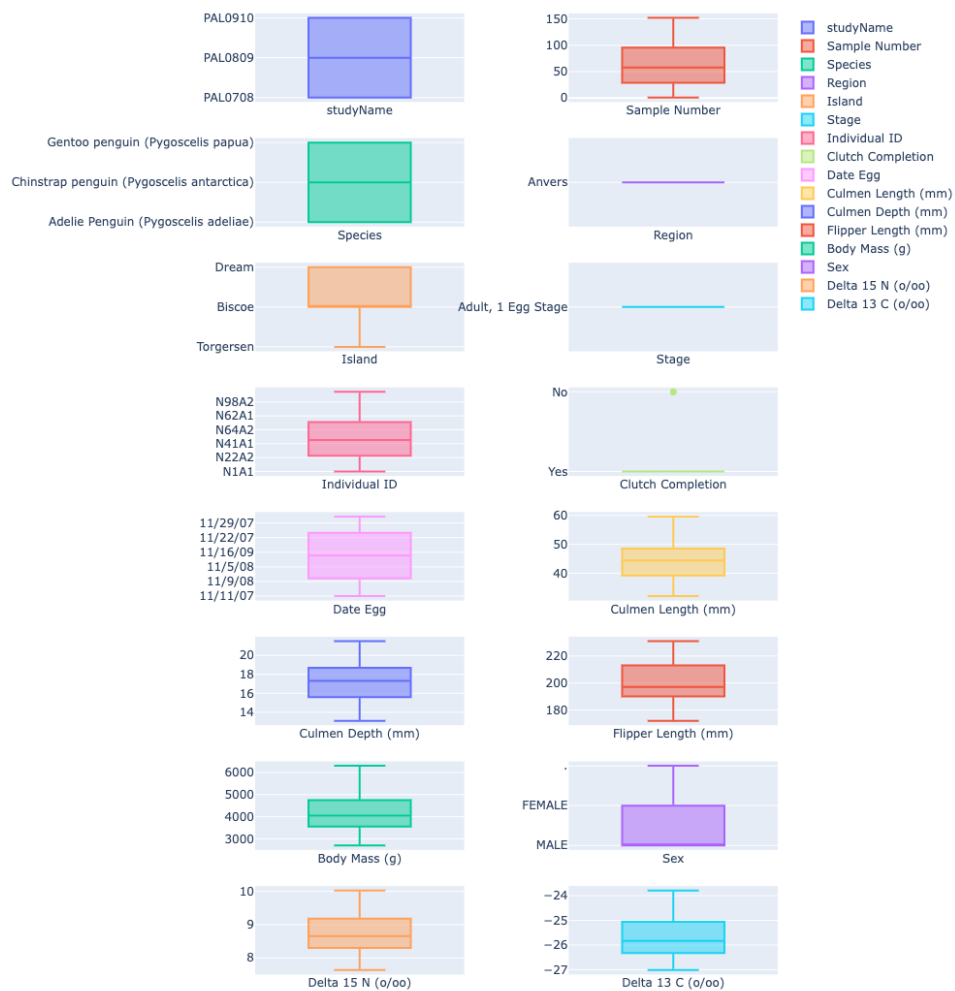
Vycházeli jsme z předpokladu, že odlehlé hodnoty jsou pouze ty, nacházející se naprosto mimo rozsah datové sady, tak jako to bylo bráno na přednáškách, viz obrázek z prezentace 3.1.



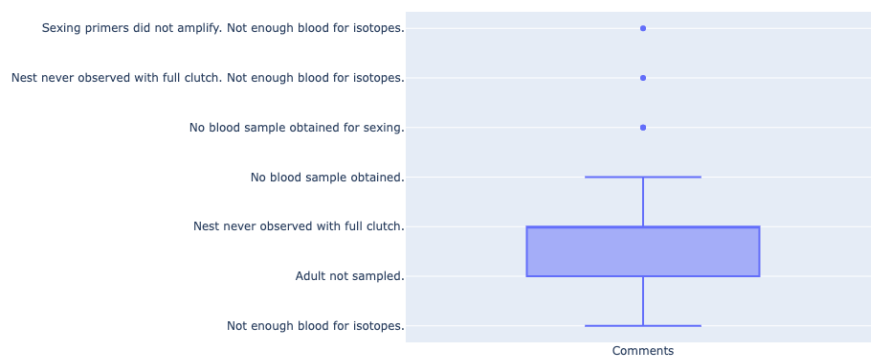
Obrázek 3.1: Zdroj: Materiály k democvičení č. 1 Porozumění datům a příprava dat, dostupné z IS VUT.

Abychom se přiblížili výše uvedené vizualizaci, využili jsme tzv. boxplot, který umožňuje zobrazit rozložení hodnot s využitím percentilů. Z vytvořených grafů (viz obrázky 3.2 3.3) je vidět, že v datové sadě se nenachází prakticky žádný atribut s odlehlými hodnotami. Ty se nachází pouze u atributu *Comments*, *Clutch Completion* a *Sex* v podobě tečky

Detekce odlehlých hodnot pomocí boxplotu



Obrázek 3.2: Vizualizace rozložení hodnot pomocí boxplotů.



Obrázek 3.3: Vizualizace rozložení atributu Comments pomocí boxplotu.

Kapitola 4

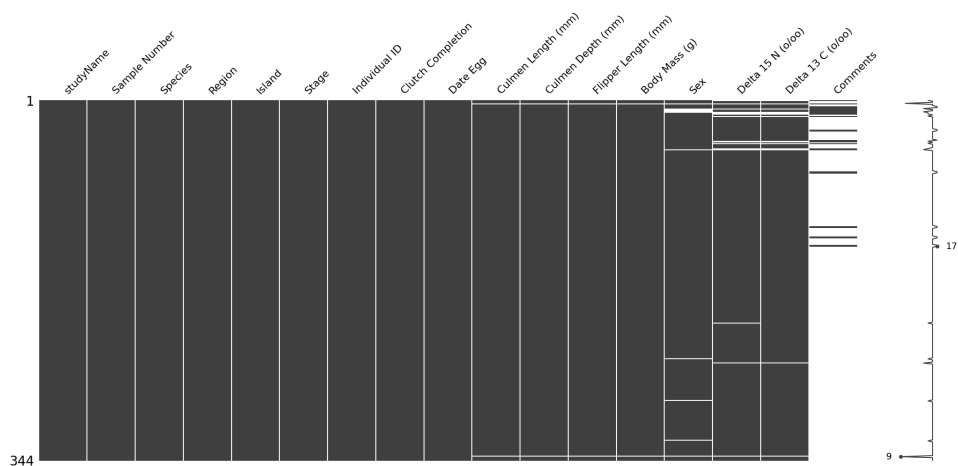
Analýza chybějících hodnot

Analýza chybějících hodnot měla za následek zjištění počtu a procentuálního zastoupení chybějících hodnot napříč atributy. Jejich přesné hodnoty jsou k dispozici v tabulce 4.1 a na obrázku 4.1 je dostupná jejich vizuální reprezentace. Celkově chybí 363 hodnot v datové sadě a 17 objektů má více než jednu chybějící hodnotu.

Z tabulky je zřejmé, že většina atributů má vyplněné všechny hodnoty. Pod prahem hodnoty 1 % se vyskytují pouze atributy *Culmen Length (mm)*, *Culmen Depth (mm)*, *Flipper Length (mm)* a *Body Mass (g)*. Znatelnější počet chybějících hodnot obsahují atributy *Sex*, *Delta 13 C (o/oo)* a *Delta 15 N (o/oo)*, které mají pod 5 % chybějících hodnot. Atribut *Comments* obsahuje největší procentuální zastoupení chybějících hodnot a to vysoké hodnoty 92.44 %. Tato vysoká hodnota zamezuje efektivnímu vyplnění chybějících hodnot nějakou jinou hodnotou (například průměrem).

| Název atributu | Počet chybějících hodnot | Počet chybějících hodnot [%] |
|---------------------|--------------------------|------------------------------|
| studyName | 0 | 0.0 |
| Sample Number | 0 | 0.0 |
| Species | 0 | 0.0 |
| Region | 0 | 0.0 |
| Island | 0 | 0.0 |
| Stage | 0 | 0.0 |
| Individual ID | 0 | 0.0 |
| Clutch Completion | 0 | 0.0 |
| Date Egg | 0 | 0.0 |
| Culmen Length (mm) | 2 | 0.58 |
| Culmen Depth (mm) | 2 | 0.58 |
| Flipper Length (mm) | 2 | 0.58 |
| Body Mass (g) | 2 | 0.58 |
| Sex | 10 | 2.91 |
| Delta 15 N (o/oo) | 14 | 4.07 |
| Delta 13 C (o/oo) | 13 | 3.78 |
| Comments | 318 | 92.44 |
| | 363 | 6.62 |

Tabulka 4.1: Souhrn chybějících hodnot atributů.



Obrázek 4.1: Vizualizace chybějících hodnot atributů, kde bílá barva znázorňuje chybějící hodnoty.

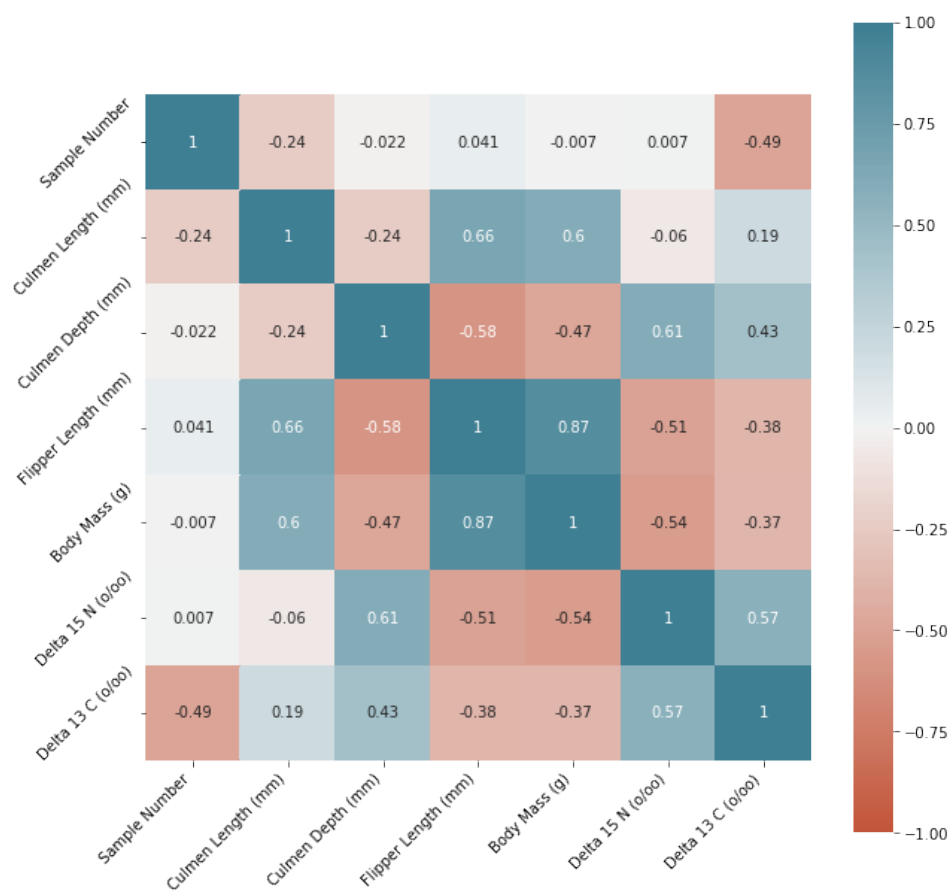
Kapitola 5

Korelační analýza numerických atributů

V prvé řadě je zde třeba vysvětlit samotný pojem korelační analýza a s ní související korelační koeficient. Samotná korelace znázorňuje vzájemný vztah dvou veličin. Dvě proměnné jsou korelované, jestliže určité hodnoty jedné proměnné mají tendenci se vyskytovat společně s určitými hodnotami druhé proměnné. Korelační koeficient je pak číslo, které znázorňuje sílu tohoto vztahu.

K zobrazení korelace jednotlivých atributů jsme využili teplotní mapu (viz obrázek 5.1), která za pomoci barev zobrazuje sílu korelace daných proměnných – modrá zobrazuje plnou korelaci a velmi nízkou, takřka žádnou. Modrá je zpravidla diagonála, tedy prvek je v korelaci sám se sebou. Z naší datové sady dále vyplývá, že se sebou korelují také atributy *Body mass* a *Fliper Lenght (mm)*, tedy spolu mají vzájemnou souvislost tělesná váha a délka ploutve tučňáka. Jistou korelaci je také možné pozorovat i mezi tělesnou vahou a atributem *Culmen Lenght*, který vyjadřuje délku zobáku, z toho je jistě jasná i korelace mezi délkou zobáku a délkou ploutve.

KAPITOLA 5. KORELAČNÍ ANALÝZA NUMERICKÝCH ATRIBUTŮ²²



Obrázek 5.1: Korelační analýza numerických atributů

Část II

Datové sady vhodné pro dolovací algoritmy

Kapitola 6

Odstranění irelevantních atributů

Jako dolovací úlohu jsme si pro druhou část projektu vybrali klasifikaci druhů tučňáků na základě ostatních atributů, uvedenou přímo pro naši datovou sadu. Kvůli možnosti využití supervised learning jsme se rozhodli ponechat atribut *Species*, díky kterému se může algoritmus učit na již známých datech. Odstraněnými atributy jsou tedy *studyName*, *Clutch Completion*, *Date Egg*, *Individual ID*, *Sample Number* a *Comments*, které nemají žádnou souvislost s druhem daného tučňáka, ale spíše souvisí s analýzou, která byla na tučňácích prováděna. Dále byly odstraněny také *Region* a *Stage*, které disponovaly pouze jednou hodnotou, na klasifikaci by tedy neměly žádný vliv. Výsledná podoba datové sady je znázorněna výčtem prvních pěti prvků v tabulce níže.

| | Species | Island | Culmen Length (mm) | Culmen Depth (mm) | Flipper Length (mm) | Body Mass (g) | Sex | Delta 15 N (o/oo) | Delta 13 C (o/oo) |
|---|-------------------------------------|-----------|--------------------|-------------------|---------------------|---------------|--------|-------------------|-------------------|
| 0 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | 39.1 | 18.7 | 181.0 | 3750.0 | MALE | | |
| 1 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | 39.5 | 17.4 | 186.0 | 3800.0 | FEMALE | 8.94956 | -24.69454 |
| 2 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | 40.3 | 18.0 | 195.0 | 3250.0 | FEMALE | 8.36821 | -25.33302 |
| 3 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | | | | | | | |
| 4 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | 36.7 | 19.3 | 193.0 | 3450.0 | FEMALE | 8.76651 | -25.32426 |

Kapitola 7

Zpracování chybějících hodnot

Z kapitoly 4 víme, že největším počtem chybějících hodnot disponoval atribut *Comments*, a dále *Delta 13 C (o/oo)*, *Delta 15 N (o/oo)* a *Sex*. V počtech jednotek pak chyběly hodnoty i u atributů *Culmen Length (mm)*, *Culmen Depth (mm)*, *Flipper Length (mm)* a *Body Mass (g)*. Atribut *Comments* byl odstraněn již v rámci nepotřebných atributů, dále bylo tedy potřeba se vypořádat se zbývajících. Protože atribut *Sex* by se dal jen těžko nahradit průměrnou hodnotou, a protože v aktuální fázi ještě nejsme schopni tučňáky kategorizovat – data slouží teprve pro trénink této kategorizace, řádky, ve kterých atribut pohlaví chyběl, byly z datové sady odstraněny. U číselných hodnot, které jsou v našem případě všechny ostatní zmíněné výše, byly prázdné hodnoty nahrazeny hodnotami průměrnými, které byly spočítány z ostatních dostupných hodnot datové sady.

Kapitola 8

Zpracování odlehlých hodnot

Během analýzy v kapitole 3 jsme zjistili, že v datové sadě se nenachází mnoho odlehlých hodnot. Z grafů dostupných v téže kapitole je tedy vidět, že odlehlá hodnota v podobě tečky se nachází u atributu pohlaví. Tato hodnota byla vyřešena stejně jako chybějící hodnota, daný řádek byl tedy odstraněn. Další odlehlé hodnoty se nachází u atributu *Comments*. Ten ale patří mezi irrelevantní atributy, které byly odstraněny hned v začátku zpracování datové sady, takže s těmito odlehlými hodnotami nebylo třeba nic dělat, protože se v této fázi v datové sadě již nenachází. Stejným případem byly i odlehlé hodnoty atributu *Clutch Completion*.

Kapitola 9

Datová sada číslo I.

9.1 Diskretizace numerických atributů

Diskretizace numerických hodnot byla provedena ručně pro každý atribut zvlášť. Byla využita k tomu určená funkce *cut* knihovny *pandas*. Tato funkce umožňuje datový set rozdělit na intervaly stejné velikosti, které jsou následně pojmenovány. V našem případě jsme se rozhodli data rozdělit vždy do tří intervalů a pojmenovat je následovně:

- Pro atribut Culmen Length (mm) ["short", "medium", "long"]
- Pro atribut Culmen Depth (mm) ["short", "medium", "long"]
- Pro atribut Flipper Length (mm) ["short", "medium", "long"]
- Pro atribut Body Mass (g) ["light", "middleweight", "heavy"]
- Pro atribut Delta 15 N (o/oo) ["weak", "medium concentrated", "highly concentrated"]
- Pro atribut Delta 13 C (o/oo) ["weak", "medium concentrated", "highly concentrated"]

V následující tabulce je prvních pět řádků transformovaného datového setu, na kterém můžeme pozorovat ukázkou výsledné datové sady vhodné pro algoritmy, které vyžadují na vstupu kategorické atributy.

| | Species | Island | Culmen Length (mm) | Culmen Depth (mm) | Flipper Length (mm) | Body Mass (g) | Sex | Delta 15 N (o/oo) | Delta 13 C (o/oo) |
|---|--|-----------|--------------------------|-------------------------|---------------------------|---------------------|--------|----------------------|----------------------|
| 0 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | short | medium | short | light | MALE | medium concentrated | medium concentrated |
| 1 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | short | medium | short | light | FEMALE | medium concentrated | highly concentrated |
| 2 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | short | medium | medium | light | FEMALE | weak | medium concentrated |
| 4 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | short | long | medium | light | FEMALE | medium concentrated | medium concentrated |
| 5 | Adelie Penguin (Pygoscelis adeliae) | Torgersen | short | long | short | light | MALE | medium concentrated | medium concentrated |

Kapitola 10

Datová sada číslo II.

10.1 Transformace kategorických atributů na numerické a jejich následná normalizace

Transformace kategorických atributů na numerické atributy probíhala v několika krocích. Nejdříve byly ve `for` smyčce vybrány pouze kategorické atributy, jejichž názvy byly uloženy do pole výběru. S využitím `OrdinalEncoder()`, který je pro tento účel již zabudován v knihovně `sklearn`, byl následně tento výběr atributů transformován na numerické. Celá datová sada byla poté ještě normalizována na interval $<0,1>$, a to opětovně s využitím této knihovny, tentokrát však za pomoci funkce `MinMaxScaler()`. Ukázka výsledků těchto operací nad daty je vidět v tabulce níže, ve které je zobrazeno prvních pět řádků datové sady.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|-----|-----|---------------------|--------------------|---------------------|---------------------|-----|---------------------|--------------------|
| 0 | 0.0 | 1.0 | 0.25454545454545463 | 0.6666666666666665 | 0.15254237288135597 | 0.29166666666666674 | 1.0 | 0.46012171657238676 | 0.4123497569430219 |
| 1 | 0.0 | 1.0 | 0.26909090909090905 | 0.5119047619047619 | 0.2372881355932206 | 0.3055555555555556 | 0.5 | 0.5504504353930235 | 0.719310897683906 |
| 2 | 0.0 | 1.0 | 0.2981818181818181 | 0.5833333333333335 | 0.38983050847457656 | 0.15277777777777778 | 0.5 | 0.3075370627266798 | 0.5216922995973219 |
| 3 | 0.0 | 1.0 | 0.16727272727272724 | 0.7380952380952381 | 0.3559322033898309 | 0.20833333333333337 | 0.5 | 0.47396416573348255 | 0.5244036436006407 |
| 4 | 0.0 | 1.0 | 0.2618181818181817 | 0.8928571428571432 | 0.30508474576271194 | 0.26388888888888884 | 1.0 | 0.4315321488860291 | 0.5325160096196999 |