```python
In [53]: import os
         from collections import Counter
         from PIL import Image

         # Sample path to NABirds dataset
         DATASET_DIR = r"C:/Users/kevsa/Downloads/nabirds"

         def read_id_file(path):
             mapping = {}
             with open(path, "r") as f:
                 for line in f:
                     parts = line.strip().split()
                     mapping[parts[0]] = parts[1:]
             return mapping

         def main():

             print("\n--- VALIDATING NABirds DATASET ---\n")

             # Load data
             images = read_id_file(os.path.join(DATASET_DIR, "images.txt"))
             labels = read_id_file(os.path.join(DATASET_DIR, "image_class_labels.txt"
             split = read_id_file(os.path.join(DATASET_DIR, "train_test_split.txt"))
             classes = read_id_file(os.path.join(DATASET_DIR, "classes.txt"))

             print(f"Total images listed: {len(images)}")
             print(f"Total labels listed: {len(labels)}")
             print(f"Total split entries: {len(split)}")
             print(f"Total classes: {len(classes)}")

             # 1. Image Existence Check
             print("\n[1] Checking image files exist...")
             missing_images = []
             for image_id, (relative_path,) in images.items():
                 image_path = os.path.join(DATASET_DIR, "images", *relative_path.spli
                 if not os.path.exists(image_path):
                     missing_images.append(relative_path)
             print(f"Missing images: {len(missing_images)}")
             if missing_images[:5]:
                 print("Example missing:", missing_images[:5])

             # 2. Label Integrity Check

             print("\n[2] Checking label coverage...")
             missing_labels = []
             for image_id in images:
                 if image_id not in labels:
                     missing_labels.append(image_id)
             print(f"Images missing labels: {len(missing_labels)}")

             # 3. Train/Test Split Check

             print("\n[3] Validating train/test split...")
             train_ids = {iid for iid, (flag,) in split.items() if flag == '1'}
```

```python
    test_ids = {iid for iid, (flag,) in split.items() if flag == '0'}
    overlap = train_ids.intersection(test_ids)
    print(f"Train images: {len(train_ids)}")
    print(f"Test images: {len(test_ids)}")
    print(f"Overlap: {len(overlap)}")

    # 4. Class Distribution Check

    print("\n[4] Checking class distribution...")
    class_counts = Counter()
    for image_id, (class_id,) in labels.items():
        class_counts[int(class_id)] += 1
    print(f"Unique classes in labels: {len(class_counts)}")
    print(f"Largest class count: {max(class_counts.values())}")
    print(f"Smallest class count: {min(class_counts.values())}")

    # 5. Image Load & Size Check

    print("\n[5] Testing image loading (sample 200)...")
    bad_images = []
    for i, (image_id, (relative_path,)) in enumerate(images.items()):
        image_path = os.path.join(DATASET_DIR, "images", *relative_path.spli
        try:
            with Image.open(image_path) as img:
                img.verify()
        except Exception as e:
            bad_images.append((relative_path, str(e)))
        if i >= 200:  # sample subset for speed
            break
    print(f"Corrupt/unreadable images (sample): {len(bad_images)}")
    if bad_images[:3]:
        print("Example issues:", bad_images[:3])

    # 6. Bounding Box Check

    bbox_path = os.path.join(DATASET_DIR, "bounding_boxes.txt")
    if os.path.exists(bbox_path):
        print("\n[6] Validating bounding boxes...")
        bboxes = read_id_file(bbox_path)
        invalid_boxes = []
        for image_id, vals in bboxes.items():
            x, y, w, h = map(float, vals)
            if w <= 0 or h <= 0:
                invalid_boxes.append(image_id)
        print(f"Invalid bounding boxes: {len(invalid_boxes)}")

    # 7. Part Locations Check

    parts_loc_path = os.path.join(DATASET_DIR, "parts", "part_locs.txt")
    if os.path.exists(parts_loc_path):
        print("\n[7] Validating part locations...")
        invalid_parts = 0
        with open(parts_loc_path, "r") as f:
            for line in f:
                image_id, part_id, x, y, visible = line.strip().split()
                x, y = float(x), float(y)
```

```
                if x < 0 or y < 0:
                    invalid_parts += 1
        print(f"Invalid part coordinates: {invalid_parts}")

    if __name__ == "__main__":
        main()
```

```
--- VALIDATING NABirds DATASET ---

Total images listed: 48562
Total labels listed: 48562
Total split entries: 48562
Total classes: 1011

[1] Checking image files exist...
Missing images: 0

[2] Checking label coverage...
Images missing labels: 0

[3] Validating train/test split...
Train images: 23929
Test images: 24633
Overlap: 0

[4] Checking class distribution...
Unique classes in labels: 555
Largest class count: 120
Smallest class count: 13

[5] Testing image loading (sample 200)...
Corrupt/unreadable images (sample): 0

[6] Validating bounding boxes...
Invalid bounding boxes: 0

[7] Validating part locations...
Invalid part coordinates: 0
```

In [54]:
```python
# Images per Species
# Paths
DATASET_DIR = r"C:/Users/kevsa/Downloads/nabirds"

# Helper to read ID files
def read_id_file(path):

    mapping = {}
    with open(path, "r") as f:
        for line in f:
            parts = line.strip().split()
            mapping[parts[0]] = parts[1:]
    return mapping

def count_images_per_class():
    # Load images and labels
    images = read_id_file(os.path.join(DATASET_DIR, "images.txt"))
```

Loading [MathJax]/extensions/Safe.js

```python
    labels = read_id_file(os.path.join(DATASET_DIR, "image_class_labels.txt"
    classes = read_id_file(os.path.join(DATASET_DIR, "classes.txt"))

    # Count number of images per class
    class_counts = Counter()
    for image_id, (class_id,) in labels.items():
        class_counts[class_id] += 1

    # Sort classes by count descending
    sorted_classes = sorted(class_counts.items(), key=lambda x: x[1], revers

    # Print results
    print(f"{'Class ID':<10} {'Class Name':<30} {'Image Count':<10}")

    for class_id, count in sorted_classes:
        class_name = " ".join(classes.get(class_id, ["Unknown"]))
        print(f"{class_id:<10} {class_name:<30} {count:<10}")

if __name__ == "__main__":
    count_images_per_class()
```

Class ID	Class Name	Image Count
900	Savannah Sparrow	120
929	Eastern Phoebe	120
700	Red-tailed Hawk (Light morph immature)	120
947	Red-eyed Vireo	120
949	Steller's Jay	120
504	Killdeer	120
766	White-crowned Sparrow (Adult)	120
840	Ruby-crowned Kinglet	120
888	Spotted Towhee	120
469	Wild Turkey	120
910	Western Meadowlark	120
477	Anhinga	120
299	Red-tailed Hawk (Light morph adult)	120
851	Gray Catbird	120
790	House Finch (Adult Male)	120
511	Willet	120
979	Northern Cardinal (Female/Juvenile)	120
835	Marsh Wren	120
756	Common Yellowthroat (Adult Male)	120
883	Black-throated Green Warbler	120
543	Barred Owl	120
952	Western Scrub-Jay	120
937	Loggerhead Shrike	120
783	Brown-headed Cowbird (Male)	120
501	American Coot	120
451	Forster's Tern	120
864	Black-and-white Warbler	120
823	Red-breasted Nuthatch	120
902	Song Sparrow	120
842	Eastern Bluebird	120
885	Wilson's Warbler	120
355	Bald Eagle (Adult, subadult)	120
746	Dark-eyed Junco (Slate-colored)	120
830	Carolina Wren	120
655	Northern Harrier (Female, immature)	120
777	Indigo Bunting (Adult Male)	120
521	Least Sandpiper	120
852	Northern Mockingbird	120
794	American Goldfinch (Breeding Male)	120
672	Ruby-throated Hummingbird (Female, immature)	120
642	Common Loon (Nonbreeding/juvenile)	120
935	Eastern Kingbird	120
397	Ring-billed Gull (Adult)	120
950	Blue Jay	120
350	Little Blue Heron (Adult)	120
320	Northern Shoveler (Breeding male)	120
314	Wood Duck (Breeding male)	120
769	Chipping Sparrow (Breeding)	120
848	Hermit Thrush	120
875	Yellow Warbler	120
836	Blue-gray Gnatcatcher	120
824	White-breasted Nuthatch	120
879	Prairie Warbler	120
363	American Kestrel (Adult male)	120
643	Horned Grebe (Nonbreeding/juvenile)	120

873	Northern Parula	120
450	Common Tern	120
822	Bushtit	120
522	Short-billed Dowitcher	120
483	Tricolored Heron	120
529	Mourning Dove	120
657	Cooper's Hawk (Immature)	120
485	Green Heron	120
512	Lesser Yellowlegs	120
1001	American Goldfinch (Female/Nonbreeding Male) 120
559	Downy Woodpecker	120
928	Black Phoebe	120
819	Tufted Titmouse	120
773	Rose-breasted Grosbeak (Adult Male) 120
317	Mallard (Breeding male)	120
914	Great-tailed Grackle	120
553	Red-bellied Woodpecker	120
938	Northern Shrike	120
987	Red-winged Blackbird (Female/juvenile) 120
772	Northern Cardinal (Adult Male) 120
352	Black-crowned Night-Heron (Adult) 120
856	European Starling (Nonbreeding Adult) 120
989	Brewer's Blackbird (Male)	120
881	Townsend's Warbler	120
753	American Robin (Adult)	119
764	White-throated Sparrow (White-striped) 119
780	Red-winged Blackbird (Male)	119
926	Pacific-slope Flycatcher	119
395	Laughing Gull (Breeding)	119
868	Nashville Warbler	119
796	House Sparrow (Male)	119
973	White-crowned Sparrow (Immature) 119
793	Lesser Goldfinch (Adult Male)	119
681	Northern Flicker (Red-shafted) 119
554	Yellow-bellied Sapsucker	119
673	Black-chinned Hummingbird (Female, immature) 119
762	Black-throated Blue Warbler (Adult Male) 119
382	Northern Flicker (Yellow-shafted) 119
778	Painted Bunting (Adult Male)	119
381	Red-headed Woodpecker (Adult)	119
561	Pileated Woodpecker	119
964	American Redstart (Female/juvenile) 119
811	Carolina Chickadee	118
857	American Pipit	118
912	Common Grackle	118
884	Canada Warbler	118
499	Merlin	118
669	Dunlin (Nonbreeding/juvenile)	118
616	Mallard (Female/Eclipse male)	118
560	Hairy Woodpecker	118
354	White Ibis (Adult)	118
831	Bewick's Wren	118
357	Sharp-shinned Hawk (Adult)	118
446	Bonaparte's Gull	118
930	Say's Phoebe	118
958	Yellow-rumped Warbler (Breeding Audubon's) 118

400	Herring Gull (Adult)	117
528	White-winged Dove	117
358	Cooper's Hawk (Adult)	117
458	Mute Swan	117
827	Brown Creeper	117
527	Eurasian Collared-Dove	117
1003	House Sparrow (Female/Juvenile)	117
957	American Crow	117
963	Common Yellowthroat (Female/immature male)	116
760	Chestnut-sided Warbler (Breeding male)	116
903	Lincoln's Sparrow	116
850	Varied Thrush	116
338	California Quail (Male)	116
908	Dickcissel	116
747	Yellow-rumped Warbler (Breeding Myrtle)	116
804	Tree Swallow	116
509	Solitary Sandpiper	116
315	Gadwall (Breeding male)	116
802	Common Raven	116
866	Tennessee Warbler	116
748	European Starling (Breeding Adult)	116
515	Marbled Godwit	116
470	Pied-billed Grebe	116
779	Bobolink (Breeding male)	115
960	American Robin (Juvenile)	115
898	Vesper Sparrow	115
886	Yellow-breasted Chat	115
832	House Wren	114
814	Chestnut-backed Chickadee	114
913	Boat-tailed Grackle	114
867	Orange-crowned Warbler	114
797	Vermilion Flycatcher (Adult male)	113
932	Great Crested Flycatcher	113
798	Yellow-rumped Warbler (Winter/juvenile Myrtle)	113
925	Least Flycatcher	113
974	Golden-crowned Sparrow (Immature)	113
904	Swamp Sparrow	113
510	Greater Yellowlegs	113
803	Horned Lark	113
375	Anna's Hummingbird (Adult Male)	113
997	House Finch (Female/immature)	112
498	Crested Caracara	112
899	Lark Sparrow	112
946	Warbling Vireo	112
968	Blackpoll Warbler (Female/juvenile)	112
481	Great Egret	111
862	Northern Waterthrush	111
775	Blue Grosbeak (Adult Male)	111
853	Brown Thrasher	111
749	Fox Sparrow (Red)	111
874	Blackburnian Warbler	111
503	Semipalmated Plover	111
351	Reddish Egret (Dark morph)	110
523	Wilson's Snipe	110
894	American Tree Sparrow	110
890	Eastern Towhee	110

```
662      American Kestrel (Female, immature) 110
656      Sharp-shinned Hawk (Immature)   109
536      Barn Owl                        109
549      Belted Kingfisher               109
348      Double-crested Cormorant (Immature) 109
934      Western Kingbird                109
674      Anna's Hummingbird (Female, immature) 109
871      Hooded Warbler                  109
813      Mountain Chickadee              109
786      Baltimore Oriole (Adult male)   109
895      Clay-colored Sparrow            109
774      Black-headed Grosbeak (Adult Male) 109
922      Pine Siskin                     108
368      Spotted Sandpiper (Breeding)    108
490      Turkey Vulture                  108
924      Eastern Wood-Pewee              108
676      Calliope Hummingbird (Female, immature) 108
516      Ruddy Turnstone                 108
484      Cattle Egret                    108
877      Pine Warbler                    108
380      Allen's Hummingbird (Adult Male) 107
795      Evening Grosbeak (Adult Male)   107
847      Swainson's Thrush               107
335      Hooded Merganser (Breeding male) 106
810      Barn Swallow                    106
491      Osprey                          106
319      Cinnamon Teal (Male)            106
347      Northern Gannet (Adult, Subadult) 106
843      Western Bluebird                106
332      Bufflehead (Breeding male)      106
333      Common Goldeneye (Breeding male) 105
525      Rock Pigeon                     105
923      Western Wood-Pewee              105
965      Magnolia Warbler (Female/immature male) 105
519      Semipalmated Sandpiper          105
962      Snow Bunting (Nonbreeding)      104
834      Winter Wren                     104
876      Palm Warbler                    104
473      Wood Stork                      104
482      Snowy Egret                     104
860      Ovenbird                        103
757      American Redstart (Adult Male)  103
541      Northern Pygmy-Owl              103
539      Great Horned Owl                103
788      Purple Finch (Adult Male)       103
897      Field Sparrow                   103
878      Yellow-throated Warbler         103
654      Bald Eagle (Immature, juvenile) 103
530      Inca Dove                       103
872      Cape May Warbler                103
865      Prothonotary Warbler            103
990      Brown-headed Cowbird (Female/Juvenile) 103
776      Lazuli Bunting (Adult Male)     102
480      Great Blue Heron                102
508      American Avocet                 102
374      Black-chinned Hummingbird (Adult Male) 101
```

980	Rose-breasted Grosbeak (Female/immature male)	101
647	Double-crested Cormorant (Adult)	101
844	Mountain Bluebird	101
366	Common Gallinule (Adult)	101
607	Herring Gull (Immature)	101
513	Whimbrel	101
373	Ruby-throated Hummingbird (Adult Male)	100
944	Blue-headed Vireo	100
651	Black-crowned Night-Heron (Immature)	99
557	Ladder-backed Woodpecker	99
550	Acorn Woodpecker	99
542	Burrowing Owl	99
758	Magnolia Warbler (Breeding male)	99
555	Red-naped Sapsucker	98
667	Spotted Sandpiper (Nonbreeding/juvenile)	98
768	Dark-eyed Junco (Oregon)	98
353	Yellow-crowned Night-Heron (Adult)	98
955	Black-billed Magpie	98
486	Glossy Ibis	98
976	Chipping Sparrow (Immature/nonbreeding adult)	98
993	Baltimore Oriole (Female/Immature male)	98
785	Bullock's Oriole (Adult male)	98
544	Northern Saw-whet Owl	98
489	Black Vulture	97
770	Scarlet Tanager (Breeding Male)	97
859	Cedar Waxwing	97
322	Green-winged Teal (Male)	97
551	Gila Woodpecker	97
948	Gray Jay	96
316	American Wigeon (Breeding male)	96
613	Wood Duck (Female/Eclipse male)	96
514	Long-billed Curlew	96
812	Black-capped Chickadee	96
356	Northern Harrier (Adult male)	95
495	Harris's Hawk	95
668	Sanderling (Nonbreeding/juvenile)	95
533	Yellow-billed Cuckoo	94
771	Western Tanager (Breeding Male)	94
640	Red-throated Loon (Nonbreeding/juvenile)	94
939	White-eyed Vireo	94
538	Eastern Screech-Owl	94
343	Common Loon (Breeding)	94
849	Wood Thrush	93
379	Rufous Hummingbird (Adult Male)	93
784	Hooded Oriole (Adult male)	93
448	Caspian Tern	93
340	Ring-necked Pheasant (Male)	93
1002	Evening Grosbeak (Female/Juvenile)	93
453	Black Skimmer	92
520	Western Sandpiper	92
892	California Towhee	92
558	Nuttall's Woodpecker	92
328	Harlequin Duck (Male)	92
750	Summer Tanager (Adult Male)	91
995	Purple Finch (Female/immature)	91
781	Yellow-headed Blackbird (Adult Male)	91

761	Blackpoll Warbler (Breeding male) 90
907	Pyrrhuloxia	90
371	Wilson's Phalarope (Breeding)	90
911	Rusty Blackbird	90
845	Townsend's Solitaire	89
323	Canvasback (Breeding male)	89
372	Broad-billed Hummingbird (Adult Male) 89
954	Clark's Nutcracker	89
981	Black-headed Grosbeak (Female/immature male) 89
767	Golden-crowned Sparrow (Adult) 88
751	Orchard Oriole (Adult Male)	88
984	Indigo Bunting (Female/juvenile) 88
1009	Yellow-rumped Warbler (Winter/juvenile Audubon's) 88
985	Painted Bunting (Female/juvenile) 87
396	Heermann's Gull (Adult)	87
635	Common Merganser (Female/immature male) 87
846	Veery	87
649	Little Blue Heron (Immature)	87
297	Ruddy Duck (Breeding male)	87
321	Northern Pintail (Breeding male) 87
861	Louisiana Waterthrush	86
517	Black Turnstone	86
500	Prairie Falcon	86
545	Common Nighthawk	86
602	Laughing Gull (Nonbreeding/Immature) 86
660	Rough-legged Hawk (Light morph) 86
828	Cactus Wren	86
945	Hutton's Vireo	85
327	Lesser Scaup (Breeding male)	85
556	Red-breasted Sapsucker	85
891	Canyon Towhee	84
402	Great Black-backed Gull (Adult) 84
821	Verdin	83
318	Blue-winged Teal (Male)	83
971	White-throated Sparrow (Tan-striped/immature) 83
666	Black-bellied Plover (Nonbreeding/juvenile) 83
817	Oak Titmouse	82
652	Yellow-crowned Night-Heron (Immature) 82
838	American Dipper	82
931	Ash-throated Flycatcher	82
472	Clark's Grebe	82
367	Black-bellied Plover (Breeding) 82
325	Ring-necked Duck (Breeding male) 82
362	Golden Eagle (Adult)	81
393	Pigeon Guillemot (Breeding)	81
658	Red-shouldered Hawk (Immature) 81
982	Blue Grosbeak (Female/juvenile) 81
378	Broad-tailed Hummingbird (Adult Male) 81
789	Cassin's Finch (Adult Male)	80
825	Pygmy Nuthatch	80
604	Ring-billed Gull (Immature)	80
933	Cassin's Kingbird	80
505	American Oystercatcher	80
459	Trumpeter Swan	80
890	Rufous-crowned Sparrow	80
494	Mississippi Kite	80

901	Fox Sparrow (Sooty)	79
619	Northern Shoveler (Female/Eclipse male)	79
858	Bohemian Waxwing	79
854	Curve-billed Thrasher	78
967	Chestnut-sided Warbler (Female/immature male)	78
636	Red-breasted Merganser (Female/immature male)	78
896	Brewer's Sparrow	78
401	Glaucous-winged Gull (Adult)	78
377	Calliope Hummingbird (Adult Male)	78
364	Peregrine Falcon (Adult)	77
552	Golden-fronted Woodpecker	77
806	Northern Rough-winged Swallow	77
496	Swainson's Hawk (Light morph)	77
936	Scissor-tailed Flycatcher	77
870	Mourning Warbler	77
337	Red-breasted Merganser (Breeding male)	77
698	Ruddy Duck (Female/juvenile)	76
994	Pine Grosbeak (Female/juvenile)	76
399	California Gull (Adult)	76
452	Royal Tern	76
909	Eastern Meadowlark	76
370	Dunlin (Breeding)	76
344	Horned Grebe (Breeding)	76
1004	Vermilion Flycatcher (Female, immature)	76
349	Great Cormorant (Adult)	75
526	Band-tailed Pigeon	75
880	Black-throated Gray Warbler	75
468	Ruffed Grouse	75
920	Common Redpoll	75
471	Western Grebe	75
805	Violet-green Swallow	75
942	Plumbeous Vireo	75
855	California Thrasher	75
531	Common Ground-Dove	75
537	Western Screech-Owl	74
376	Costa's Hummingbird (Adult Male)	74
507	Black-necked Stilt	74
839	Golden-crowned Kinglet	74
808	Cliff Swallow	74
863	Blue-winged Warbler	74
365	Purple Gallinule (Adult)	74
943	Cassin's Vireo	73
1000	Lesser Goldfinch (Female/juvenile)	73
346	Eared Grebe (Breeding)	73
988	Yellow-headed Blackbird (Female/Immature Male)	72
1005	European Starling (Juvenile)	72
493	White-tailed Kite	72
680	Red-headed Woodpecker (Immature)	72
540	Snowy Owl	72
969	Black-throated Blue Warbler (Female/Immature male)	72
326	Greater Scaup (Breeding male)	72
826	Brown-headed Nuthatch	72
1010	Dark-eyed Junco (Red-backed/Gray-headed)	72
534	Black-billed Cuckoo	71
961	Phainopepla (Male)	71
454	Greater White-fronted Goose	71

915	Bronzed Cowbird	71
339	Gambel's Quail (Male)	71
927	Cordilleran Flycatcher	70
679	Allen's Hummingbird (Female, immature)	70
398	Western Gull (Adult)	70
359	Red-shouldered Hawk (Adult)	70
678	Rufous Hummingbird (Female, immature)	70
329	Surf Scoter (Male)	70
752	Purple Martin (Adult male)	70
820	Black-crested Titmouse	70
941	Yellow-throated Vireo	69
991	Hooded Oriole (Female/Immature male)	69
986	Bobolink (Female/juvenile/nonbreeding male)	69
998	Red Crossbill (Female/juvenile)	67
970	Lark Bunting (Female/Nonbreeding male)	67
298	Swainson's Hawk (Dark morph)	67
959	Purple Martin (Female/juvenile)	67
972	Harris's Sparrow (Immature)	67
324	Redhead (Breeding male)	66
782	Brewer's Blackbird (Female/Juvenile)	66
882	Hermit Warbler	66
869	MacGillivray's Warbler	66
487	White-faced Ibis	66
474	Brandt's Cormorant	66
634	Hooded Merganser (Female/immature male)	66
977	Scarlet Tanager (Female/Nonbreeding Male)	66
548	White-throated Swift	66
479	Brown Pelican	65
506	Black Oystercatcher	65
535	Greater Roadrunner	65
978	Western Tanager (Female/Nonbreeding Male)	65
631	Bufflehead (Female/immature male)	65
461	American Black Duck	65
791	Red Crossbill (Adult Male)	65
475	Neotropic Cormorant	65
645	Eared Grebe (Nonbreeding/juvenile)	64
917	Gray-crowned Rosy-Finch	64
478	American White Pelican	64
829	Canyon Wren	64
893	Abert's Towhee	64
296	Long-tailed Duck (Winter male)	64
841	Wrentit	64
765	Harris's Sparrow (Adult)	64
992	Bullock's Oriole (Female/Immature male)	64
449	Black Tern	63
518	Surfbird	63
609	Great Black-backed Gull (Immature)	63
653	White Ibis (Immature)	63
336	Common Merganser (Breeding male)	63
644	Red-necked Grebe (Nonbreeding/juvenile)	62
809	Cave Swallow	62
887	Green-tailed Towhee	62
476	Pelagic Cormorant	62
626	Lesser Scaup (Female/Eclipse male)	62
906	Summer Tanager (Female)	61
800	Fish Crow	61

457	Canada Goose	61
953	Mexican Jay	61
462	Mottled Duck	60
921	Hoary Redpoll	60
966	Bay-breasted Warbler (Female, Nonbreeding male, Immature)	60
488	Roseate Spoonbill	60
816	Bridled Titmouse	59
956	Yellow-billed Magpie	59
670	Wilson's Phalarope (Nonbreeding, juvenile)	59
905	Dark-eyed Junco (Pink-sided)	59
918	Black Rosy-Finch	59
467	Northern Bobwhite	59
632	Common Goldeneye (Female/Eclipse male)	58
999	White-winged Crossbill (Female/juvenile)	58
497	Red-tailed Hawk (Dark morph)	58
360	Broad-winged Hawk (Adult)	58
546	Chimney Swift	57
940	Bell's Vireo	57
787	Pine Grosbeak (Adult Male)	57
661	Golden Eagle (Immature)	56
677	Broad-tailed Hummingbird (Female, immature)	56
951	Florida Scrub-Jay	55
1008	Orchard Oriole (Female/Juvenile)	55
447	Mew Gull	55
460	Tundra Swan	55
996	Cassin's Finch (Female/immature)	55
605	Western Gull (Immature)	54
837	Black-tailed Gnatcatcher	54
524	American Woodcock	54
601	Black-legged Kittiwake (Immature)	54
345	Red-necked Grebe (Breeding)	54
606	California Gull (Immature)	54
641	Pacific Loon (Nonbreeding/juvenile)	54
492	Swallow-tailed Kite	53
295	Common Eider (Adult male)	53
663	Peregrine Falcon (Immature)	53
792	White-winged Crossbill (Adult Male)	53
547	Vaux's Swift	53
394	Black-legged Kittiwake (Adult)	53
455	Brant	52
610	Black-bellied Whistling-Duck	51
637	California Quail (Female/juvenile)	51
361	Rough-legged Hawk (Dark morph)	50
502	Sandhill Crane	50
919	Brown-capped Rosy-Finch	50
624	Ring-necked Duck (Female/Eclipse male)	50
916	Orchard Oriole (Immature Male)	49
608	Glaucous-winged Gull (Immature)	49
611	Snow Goose (Blue morph)	49
801	Chihuahuan Raven	49
629	White-winged Scoter (Female/juvenile)	48
815	Boreal Chickadee	48
465	Ruddy Duck (Winter male)	48
759	Bay-breasted Warbler (Breeding male)	48
620	Northern Pintail (Female/Eclipse male)	47
603	Heermann's Gull (Immature)	47

```
818       Juniper Titmouse                    47
659       Broad-winged Hawk (Immature)   47
313       Snow Goose (White morph)       47
621       Green-winged Teal (Female/juvenile) 47
612       Ross's Goose                   46
650       Reddish Egret (White morph)    46
532       Monk Parakeet                  46
696       Common Eider (Immature/Eclipse male) 45
392       Black Guillemot (Breeding)     45
334       Barrow's Goldeneye (Breeding male) 45
628       Surf Scoter (Female/immature)  44
697       Long-tailed Duck (Female/juvenile) 44
699       Swainson's Hawk (Immature)     44
983       Lazuli Bunting (Female/juvenile) 44
330       White-winged Scoter (Male)     44
648       Great Cormorant (Immature)     42
675       Costa's Hummingbird (Female, immature) 42
671       Broad-billed Hummingbird (Female, immature) 42
622       Canvasback (Female/Eclipse male) 41
463       Common Eider (Female/juvenile) 41
331       Black Scoter (Male)            41
615       American Wigeon (Female/Eclipse male) 41
630       Black Scoter (Female/juvenile) 40
466       Scaled Quail                   39
625       Greater Scaup (Female/Eclipse male) 39
646       Northern Gannet (Immature/Juvenile) 39
614       Gadwall (Female/Eclipse male)  39
833       Pacific Wren                   37
600       Pigeon Guillemot (Nonbreeding, juvenile) 37
1007      Summer Tanager (Immature Male) 37
638       Gambel's Quail (Female/juvenile) 36
799       Northwestern Crow              36
754       Phainopepla (Female/juvenile)  35
807       Bank Swallow                   35
623       Redhead (Female/Eclipse male)  34
599       Black Guillemot (Nonbreeding, juvenile) 33
617       Blue-winged Teal (Female/juvenile) 33
633       Barrow's Goldeneye (Female/Eclipse male) 32
618       Cinnamon Teal (Female/juvenile) 32
639       Ring-necked Pheasant (Female/juvenile) 32
369       Sanderling (Breeding)          31
456       Cackling Goose                 31
665       Common Gallinule (Immature)    31
342       Pacific Loon (Breeding)        31
763       Lark Bunting (Breeding male)   30
464       Long-tailed Duck (Summer male) 28
627       Harlequin Duck (Female/juvenile) 27
1006      Fox Sparrow (Thick-billed/Slate-colored) 25
755       Snow Bunting (Breeding adult)  24
341       Red-throated Loon (Breeding)   23
664       Purple Gallinule (Immature)    21
975       Dark-eyed Junco (White-winged) 13
```

In [55]: **import** os

DATASET_DIR **=** "C:/Users/kevsa/Downloads/nabirds"

Loading [MathJax]/extensions/Safe.js

```python
CLASSES_FILE = os.path.join(DATASET_DIR, "classes.txt")

def read_classes(path):
    classes = {}
    with open(path, "r") as f:
        for line in f:
            class_id, *name_parts = line.strip().split()
            class_name = " ".join(name_parts)
            classes[int(class_id)] = class_name
    return classes

def find_partial_matches(classes):
    unmatched_classes = dict(classes)
    matches = {}

    while unmatched_classes:
        # pick first item
        id1, name1 = next(iter(unmatched_classes.items()))
        key1 = name1.lower()
        group_ids = [id1]

        # compare with all other remaining classes
        to_remove = [id1]
        for id2, name2 in unmatched_classes.items():
            if id1 == id2:
                continue
            key2 = name2.lower()
            if key1 in key2 or key2 in key1:
                group_ids.append(id2)
                to_remove.append(id2)

        # add group to matches if more than 1 label
        if len(group_ids) > 1:
            matches[key1] = group_ids

        # remove processed IDs
        for rid in to_remove:
            unmatched_classes.pop(rid, None)

    return matches

def main():
    classes = read_classes(CLASSES_FILE)
    print(f"Total class labels: {len(classes)}\n")

    # Find partial matches and remove duplicates
    partial_matches = find_partial_matches(classes)

    # Sort by number of repeats, descending
    sorted_matches = sorted(partial_matches.items(), key=lambda x: len(x[1])

    print("--- Partial / repeated species labels (ordered by number of repea
    for species_name, label_ids in sorted_matches:
        print(f"{species_name}: {label_ids}")
```

```python
if __name__ == "__main__":
    main()
```

```
Total class labels: 1011

--- Partial / repeated species labels (ordered by number of repeats) ---
birds: [0, 6, 19, 22, 165, 227]
dark-eyed junco: [744, 746, 768, 905, 975, 1010]
yellow-rumped warbler: [691, 747, 798, 958, 1009]
common eider: [179, 295, 463, 696]
swainson's hawk: [202, 298, 496, 699]
long-tailed duck: [203, 296, 464, 697]
red-tailed hawk: [210, 299, 497, 700]
ruddy duck: [238, 297, 465, 698]
summer tanager: [259, 750, 906, 1007]
european starling: [439, 748, 856, 1005]
fox sparrow: [591, 749, 901, 1006]
orchard oriole: [704, 751, 916, 1008]
black-bellied plover: [23, 367, 666]
red-headed woodpecker: [31, 381, 680]
snow goose: [33, 313, 611]
northern gannet: [44, 347, 646]
spotted sandpiper: [56, 368, 667]
purple gallinule: [61, 365, 664]
california quail: [73, 338, 637]
red-throated loon: [74, 341, 640]
verdin: [80, 433, 821]
wood duck: [81, 314, 613]
horned grebe: [84, 344, 643]
gadwall: [87, 315, 614]
bushtits: [89, 434, 822]
broad-billed hummingbird: [90, 372, 671]
american wigeon: [92, 316, 615]
american kestrel: [94, 363, 662]
little blue heron: [99, 350, 649]
mallard: [102, 317, 616]
gambel's quail: [103, 339, 638]
ruby-throated hummingbird: [107, 373, 672]
bald eagle: [108, 355, 654]
common gallinule: [110, 366, 665]
sanderling: [114, 369, 668]
blue-winged teal: [116, 318, 617]
red-necked grebe: [118, 345, 644]
northern harrier: [120, 356, 655]
cinnamon teal: [122, 319, 618]
black-chinned hummingbird: [123, 374, 673]
reddish egret: [127, 351, 650]
northern shoveler: [128, 320, 619]
dunlin: [129, 370, 669]
northern pintail: [130, 321, 620]
pacific loon: [132, 342, 641]
double-crested cormorant: [134, 348, 647]
sharp-shinned hawk: [135, 357, 656]
black guillemot: [144, 392, 599]
black-legged kittiwake: [145, 394, 601]
green-winged teal: [147, 322, 621]
anna's hummingbird: [152, 375, 674]
canvasback: [153, 323, 622]
cooper's hawk: [155, 358, 657]
```

eared grebe: [157, 346, 645]
wilson's phalarope: [158, 371, 670]
redhead: [159, 324, 623]
ring-necked duck: [161, 325, 624]
costa's hummingbird: [167, 376, 675]
greater scaup: [168, 326, 625]
ring-necked pheasant: [169, 340, 639]
great cormorant: [170, 349, 648]
black-crowned night-heron: [173, 352, 651]
lesser scaup: [174, 327, 626]
red-shouldered hawk: [175, 359, 658]
peregrine falcon: [176, 364, 663]
calliope hummingbird: [180, 377, 676]
harlequin duck: [181, 328, 627]
yellow-crowned night-heron: [184, 353, 652]
broad-winged hawk: [186, 360, 659]
surf scoter: [187, 329, 628]
pigeon guillemot: [190, 393, 600]
white-winged scoter: [192, 330, 629]
common loon: [194, 343, 642]
white ibis: [198, 354, 653]
broad-tailed hummingbird: [200, 378, 677]
black scoter: [201, 331, 630]
bufflehead: [209, 332, 631]
rufous hummingbird: [213, 379, 678]
common goldeneye: [214, 333, 632]
northern flicker: [216, 382, 681]
laughing gull: [218, 395, 602]
barrow's goldeneye: [219, 334, 633]
rough-legged hawk: [224, 361, 660]
hooded merganser: [225, 335, 634]
allen's hummingbird: [228, 380, 679]
common merganser: [229, 336, 635]
heermann's gull: [232, 396, 603]
golden eagle: [234, 362, 661]
red-breasted merganser: [235, 337, 636]
western gull: [239, 398, 605]
ring-billed gull: [240, 397, 604]
california gull: [241, 399, 606]
herring gull: [242, 400, 607]
glaucous-winged gull: [243, 401, 608]
great black-backed gull: [244, 402, 609]
bobolink: [254, 779, 986]
purple martin: [264, 752, 959]
red-winged blackbird: [268, 780, 987]
scarlet tanager: [274, 770, 977]
western tanager: [291, 771, 978]
pine grosbeak: [294, 787, 994]
yellow-headed blackbird: [306, 781, 988]
northern cardinal: [310, 772, 979]
purple finch: [311, 788, 995]
common yellowthroat: [391, 756, 963]
chipping sparrow: [403, 769, 976]
cassin's finch: [408, 789, 996]
american redstart: [412, 757, 964]
brewer's blackbird: [416, 782, 989]

house finch: [419, 790, 997]
rose-breasted grosbeak: [427, 773, 980]
magnolia warbler: [430, 758, 965]
phainopepla: [441, 754, 961]
bay-breasted warbler: [443, 759, 966]
red crossbill: [444, 791, 998]
house sparrow: [445, 796, 1003]
black-headed grosbeak: [568, 774, 981]
white-winged crossbill: [574, 792, 999]
vermilion flycatcher: [575, 797, 1004]
lark bunting: [576, 763, 970]
chestnut-sided warbler: [577, 760, 967]
blackpoll warbler: [582, 761, 968]
blue grosbeak: [584, 775, 982]
black-throated blue warbler: [589, 762, 969]
lazuli bunting: [687, 776, 983]
brown-headed cowbird: [690, 783, 990]
indigo bunting: [706, 777, 984]
white-throated sparrow: [714, 764, 971]
lesser goldfinch: [715, 793, 1000]
american robin: [718, 753, 960]
hooded oriole: [719, 784, 991]
harris's sparrow: [722, 765, 972]
painted bunting: [723, 778, 985]
american goldfinch: [726, 794, 1001]
bullock's oriole: [728, 785, 992]
white-crowned sparrow: [729, 766, 973]
golden-crowned sparrow: [737, 767, 974]
evening grosbeak: [740, 795, 1002]
baltimore oriole: [742, 786, 993]
black-bellied whistling-duck: [24, 610]
semipalmated plover: [26, 503]
greater white-fronted goose: [27, 454]
black vulture: [28, 489]
american white pelican: [29, 478]
killdeer: [30, 504]
chimney swift: [34, 546]
american oystercatcher: [35, 505]
ross's goose: [36, 612]
barn owl: [37, 536]
black oystercatcher: [39, 506]
turkey vulture: [40, 490]
brown pelican: [41, 479]
brant: [42, 455]
scaled quail: [43, 466]
rock pigeon: [45, 525]
black-necked stilt: [46, 507]
acorn woodpecker: [47, 550]
pied-billed grebe: [49, 470]
american avocet: [50, 508]
cackling goose: [51, 456]
vaux's swift: [52, 547]
osprey: [53, 491]
crested caracara: [54, 498]
canada goose: [57, 457]
great blue heron: [58, 480]

solitary sandpiper: [59, 509]
mute swan: [60, 458]
western screech-owl: [62, 537]
gila woodpecker: [63, 551]
greater yellowlegs: [65, 510]
swallow-tailed kite: [66, 492]
trumpeter swan: [67, 459]
white-throated swift: [68, 548]
willet: [70, 511]
great egret: [71, 481]
tundra swan: [72, 460]
brandt's cormorant: [75, 474]
band-tailed pigeon: [76, 526]
yellow-billed cuckoo: [77, 533]
lesser yellowlegs: [78, 512]
golden-fronted woodpecker: [79, 552]
whimbrel: [82, 513]
white-tailed kite: [83, 493]
snowy egret: [85, 482]
long-billed curlew: [86, 514]
eastern screech-owl: [88, 538]
marbled godwit: [91, 515]
mississippi kite: [93, 494]
red-bellied woodpecker: [95, 553]
ruddy turnstone: [97, 516]
american black duck: [98, 461]
black turnstone: [100, 517]
neotropic cormorant: [104, 475]
eurasian collared-dove: [105, 527]
surfbird: [106, 518]
mottled duck: [109, 462]
great horned owl: [111, 539]
yellow-bellied sapsucker: [112, 554]
tricolored heron: [115, 483]
semipalmated sandpiper: [117, 519]
western sandpiper: [121, 520]
red-naped sapsucker: [124, 555]
least sandpiper: [126, 521]
northern bobwhite: [131, 467]
wood stork: [133, 473]
merlin: [136, 499]
white-winged dove: [137, 528]
monk parakeet: [138, 532]
black-billed cuckoo: [139, 534]
snowy owl: [140, 540]
common nighthawk: [141, 545]
belted kingfisher: [142, 549]
short-billed dowitcher: [146, 522]
cattle egret: [148, 484]
wilson's snipe: [149, 523]
red-breasted sapsucker: [150, 556]
american woodcock: [154, 524]
green heron: [160, 485]
american coot: [162, 501]
northern pygmy-owl: [163, 541]
ladder-backed woodpecker: [164, 557]

harris's hawk: [166, 495]
mourning dove: [171, 529]
nuttall's woodpecker: [177, 558]
burrowing owl: [182, 542]
western grebe: [185, 471]
downy woodpecker: [188, 559]
bonaparte's gull: [191, 446]
ruffed grouse: [193, 468]
pelagic cormorant: [195, 476]
inca dove: [196, 530]
greater roadrunner: [197, 535]
longspurs and snow buntings: [199, 442]
sandhill crane: [204, 502]
barred owl: [205, 543]
hairy woodpecker: [206, 560]
glossy ibis: [208, 486]
prairie falcon: [211, 500]
clark's grebe: [215, 472]
wild turkey: [220, 469]
anhinga: [221, 477]
common ground-dove: [222, 531]
white-faced ibis: [223, 487]
northern saw-whet owl: [226, 544]
pileated woodpecker: [230, 561]
roseate spoonbill: [233, 488]
mew gull: [237, 447]
caspian tern: [245, 448]
black tern: [246, 449]
common tern: [247, 450]
forster's tern: [248, 451]
royal tern: [249, 452]
black skimmer: [250, 453]
ovenbird: [251, 860]
green-tailed towhee: [252, 887]
louisiana waterthrush: [253, 861]
gray-crowned rosy-finch: [255, 917]
western wood-pewee: [256, 923]
gray jay: [257, 948]
spotted towhee: [258, 888]
northern waterthrush: [260, 862]
carolina chickadee: [261, 811]
eastern bluebird: [262, 842]
white-eyed vireo: [263, 939]
cactus wren: [265, 828]
blue-winged warbler: [266, 863]
eastern towhee: [267, 889]
black rosy-finch: [269, 918]
eastern wood-pewee: [270, 924]
steller's jay: [271, 949]
black-and-white warbler: [272, 864]
rufous-crowned sparrow: [273, 890]
gray catbird: [275, 851]
prothonotary warbler: [276, 865]
eastern meadowlark: [277, 909]
black-capped chickadee: [278, 812]
western bluebird: [279, 843]

brown-capped rosy-finch: [280, 919]
least flycatcher: [281, 925]
canyon towhee: [282, 891]
tennessee warbler: [283, 866]
red-breasted nuthatch: [284, 823]
bell's vireo: [285, 940]
blue jay: [286, 950]
tree swallow: [287, 804]
canyon wren: [288, 829]
orange-crowned warbler: [289, 867]
california towhee: [290, 892]
western meadowlark: [292, 910]
nashville warbler: [293, 868]
pacific-slope flycatcher: [300, 926]
florida scrub-jay: [301, 951]
abert's towhee: [302, 893]
mountain chickadee: [303, 813]
mountain bluebird: [304, 844]
macgillivray's warbler: [305, 869]
yellow-throated vireo: [307, 941]
mourning warbler: [308, 870]
american tree sparrow: [309, 894]
cordilleran flycatcher: [312, 927]
loggerhead shrike: [383, 937]
western scrub-jay: [384, 952]
violet-green swallow: [385, 805]
carolina wren: [386, 830]
blue-gray gnatcatcher: [387, 836]
golden-crowned kinglet: [388, 839]
northern mockingbird: [389, 852]
bohemian waxwing: [390, 858]
rusty blackbird: [404, 911]
hooded warbler: [405, 871]
chestnut-backed chickadee: [406, 814]
townsend's solitaire: [407, 845]
black phoebe: [409, 928]
clay-colored sparrow: [410, 895]
pyrrhuloxia: [411, 907]
white-breasted nuthatch: [413, 824]
plumbeous vireo: [414, 942]
mexican jay: [415, 953]
cape may warbler: [417, 872]
brewer's sparrow: [418, 896]
eastern phoebe: [420, 929]
northern rough-winged swallow: [421, 806]
bewick's wren: [422, 831]
northern parula: [423, 873]
boreal chickadee: [424, 815]
veery: [425, 846]
field sparrow: [426, 897]
clark's nutcracker: [428, 954]
common grackle: [429, 912]
vesper sparrow: [431, 898]
horned lark: [432, 803]
brown creeper: [435, 827]
american dipper: [436, 838]

wrentit: [437, 841]
brown thrasher: [438, 853]
american pipit: [440, 857]
say's phoebe: [562, 930]
cassin's vireo: [563, 943]
blackburnian warbler: [564, 874]
boat-tailed grackle: [565, 913]
black-billed magpie: [566, 955]
lark sparrow: [567, 899]
bridled titmouse: [569, 816]
swainson's thrush: [570, 847]
bank swallow: [571, 807]
house wren: [572, 832]
yellow warbler: [573, 875]
great-tailed grackle: [578, 914]
pygmy nuthatch: [579, 825]
blue-headed vireo: [580, 944]
yellow-billed magpie: [581, 956]
savannah sparrow: [583, 900]
common redpoll: [585, 920]
ash-throated flycatcher: [586, 931]
oak titmouse: [587, 817]
hermit thrush: [588, 848]
bronzed cowbird: [590, 915]
cliff swallow: [592, 808]
pacific wren: [593, 833]
black-tailed gnatcatcher: [594, 837]
ruby-crowned kinglet: [595, 840]
curve-billed thrasher: [596, 854]
cedar waxwing: [597, 859]
palm warbler: [598, 876]
northern shrike: [682, 938]
american crow: [683, 957]
great crested flycatcher: [684, 932]
hoary redpoll: [685, 921]
song sparrow: [686, 902]
pine warbler: [688, 877]
hutton's vireo: [689, 945]
juniper titmouse: [692, 818]
wood thrush: [693, 849]
lincoln's sparrow: [694, 903]
northwestern crow: [695, 799]
cassin's kingbird: [701, 933]
yellow-throated warbler: [702, 878]
pine siskin: [703, 922]
swamp sparrow: [705, 904]
cave swallow: [707, 809]
winter wren: [708, 834]
prairie warbler: [709, 879]
fish crow: [710, 800]
brown-headed nuthatch: [711, 826]
warbling vireo: [712, 946]
black-throated gray warbler: [713, 880]
western kingbird: [716, 934]
tufted titmouse: [717, 819]
california thrasher: [720, 855]

```
townsend's warbler: [721, 881]
hermit warbler: [724, 882]
chihuahuan raven: [725, 801]
eastern kingbird: [727, 935]
barn swallow: [730, 810]
marsh wren: [731, 835]
black-throated green warbler: [732, 883]
red-eyed vireo: [733, 947]
black-crested titmouse: [734, 820]
varied thrush: [735, 850]
canada warbler: [736, 884]
dickcissel: [738, 908]
common raven: [739, 802]
scissor-tailed flycatcher: [741, 936]
wilson's warbler: [743, 885]
yellow-breasted chat: [745, 886]
```