

PowerPoint Cleanup Tool - Python/Windows Guide

Remove unused images, slide masters, and layouts from PowerPoint (.pptx) files to reduce file size.

Overview

PowerPoint files often contain unused content from:

- Copied slides that brought their own masters/layouts
- Deleted slides that left behind images
- Multiple themes merged together

This tool identifies and removes that bloat using Python (works on Windows, Mac, or Linux).

Install

Python

Download and install Python 3.10+ from python.org/downloads. During installation, check "Add Python to PATH".

Verify installation:

```
python --version
```

No additional packages required - uses only Python standard library.

Usage

Step 1: Unzip the PowerPoint

PowerPoint files are ZIP archives. Choose one method:

Option A: Manual (Windows Explorer)

1. Make a copy of your .pptx file (always keep a backup!)
2. Rename the copy from .pptx to .zip
3. Right-click the .zip file → "Extract All..."
4. Extract to a folder (e.g., presentation)

Option B: PowerShell Command

```
Expand-Archive presentation.pptx -DestinationPath presentation
```

Step 2: Analyze

Open PowerShell or Command Prompt and run:

```
python ppxt_cleanup.py ./presentation/
```

This produces a report showing:

- Active vs unused masters
- Active vs unused layouts
- Active vs unused images
- Total space that can be reclaimed

Step 3: Remove Unused Content

Safe: Remove only unused images

```
python ppxt_cleanup.py ./presentation/ --remove-images
```

Moderate: Remove unused layouts (updates [Content_Types].xml)

```
python ppxt_cleanup.py ./presentation/ --remove-layouts
```

Advanced: Remove unused masters (also updates XML files)

```
python ppxt_cleanup.py ./presentation/ --remove-masters
```

Combined: Remove images and layouts together

```
python ppxt_cleanup.py ./presentation/ --remove-images --remove-layouts
```

Step 4: Re-zip as PowerPoint

Choose one method:

Option A: Manual (Windows Explorer)

1. Open the presentation folder
2. Select ALL contents inside (Ctrl+A)

3. Right-click → "Send to" → "Compressed (zipped) folder"
4. A new .zip file will be created
5. Move it outside the presentation folder
6. Rename from .zip to .pptx

IMPORTANT: Zip the CONTENTS of the folder, not the folder itself! Zipping the parent folder will cause PowerPoint to fail to open the file.

Option B: PowerShell Command

```
Compress-Archive -Path .\presentation\* -DestinationPath cleaned.zip
```

```
Rename-Item cleaned.zip cleaned.pptx
```

Step 5: Test

Open the cleaned .pptx in PowerPoint and verify all slides display correctly.

Files

File	Description
pptx_cleanup.py	Main Python script - analyzes and cleans
README_PY.md	This documentation (Python/Windows)
README_WSL.md	Documentation for Bash/WSL users

Generated Files (in presentation folder)

File	Description
remove_unused_images.sh	Script to remove unused images
remove_unused_masters.sh	Script to remove unused masters
remove_unused_layouts.sh	Script to remove unused layouts
unused_components.txt	List of all unused components

File	Description
backup_YYYYMMDD_HHMMSS/	Backup of XML files (when using --remove-masters)

What Gets Removed

Safe to Remove (--remove-images)

- Images in ppt/media/ not referenced by any active slide, master, or layout
- No XML editing required

Moderate (--remove-layouts)

- Slide layouts not used by any active slide or master
- Updates [Content_Types].xml
- Creates backup before changes
- Often the biggest source of bloat in merged presentations

Advanced (--remove-masters)

- Slide masters not used by any active slide
- Updates presentation.xml and [Content_Types].xml
- Creates backup before changes

How It Works

PowerPoint files are ZIP archives containing XML and media:

presentation.pptx (renamed .zip)

```
|—— [Content_Types].xml # File type declarations  
|—— ppt/  
|   |—— presentation.xml # Slide and master references  
|   |—— slides/      # Active slides  
|   |—— slideMasters/ # Master templates
```

```
|   |-- slideLayouts/    # Layout templates  
|   └── media/        # Images and media
```

The tool:

1. Parses presentation.xml to find active slides
 2. Traces slide → layout → master chain
 3. Identifies which media is referenced by active components
 4. Marks unreferenced content for removal
-

Troubleshooting

"presentation.xml not found" Ensure you're pointing to the unzipped folder, not the .pptx file.

PowerPoint won't open cleaned file

- Make sure you zipped the CONTENTS of the folder, not the folder itself
- Restore from backup and try removing only images first

Python not found Ensure Python is in PATH. Reinstall with "Add to PATH" checked.

Tips

- Always keep a backup of your original file
- Start by removing only images (safest)
- Test the cleaned file before deleting the original
- Large presentations with many merged themes benefit most